Towards Quality Assessment of AI Systems: A Case Study

Jesús Oviedo Lama[1,2][a], Jared David Tadeo Guerrero Sosa[2][b], Moisés Rodríguez Monje[1,2][c], Francisco Pascual Romero Chicharo[2][d] and Mario Piattini Velthuis[2][e]

[1]*AQCLab Software Quality, Ciudad Real, Spain*

[2]*Instituto Tecnologías y Sistemas de Información y Escuela Superior de Informática, Universidad de Castilla-La Mancha, Ciudad Real, Spain*

Keywords: Software Quality, Quality of Artificial Intelligence System, ISO/IEC 25000, ISO/IEC 25059.

Abstract: Artificial Intelligence is being a lever of change at all levels of society, both in public administration, in companies, organizations and even for the daily activities of individuals. Therefore, it is necessary, as in software, that Artificial Intelligence Systems obtain the results expected by the users and for this purpose, their functionality must be controlled, and their quality must be assured. This article presents the results of a functional suitability evaluation of a real Artificial Intelligence System by applying an evaluation environment based on ISO/IEC 25059 and ISO/IEC 25040 standards.

1 INTRODUCTION

Artificial Intelligence is a key part of the digital transformation that is growing both at the organisational level, where it is facilitating the robotisation and automation of processes, as well as in everyday activities: such as personal assistants, recommendations in internet searches, household appliances, etc. This growth in artificial intelligence is evidenced by objective data such as the growth in the number of artificial intelligence projects hosted on GitHub, where in the last decade the number of projects has increased from 845 in 2011 to 1.8 million in 2023. Having a significant growth of 59.3% between 2022 and 2023. It can also be seen with the number of research papers that have tripled between 2010 and 2022 (Maslej, et al., 2024).

This great impact of artificial intelligence on our society makes it necessary to generate trust in users (European Commission, 2020). For this reason, the European Union has worked to create a regulatory framework that gives rise to an 'ecosystem of trust', and the result of this work is the Artificial Intelligence Act - Regulation (EU) 2024/1689, which classifies AI systems according to their risk and establishes the controls that must be implemented. However, the Act focuses on more legal and ethical aspects and does not cover technical aspects related to the quality of AI systems.

Therefore, in order for users to achieve complete confidence in AI systems, it is necessary that they are developed with adequate quality criteria for which mature development techniques, such as software engineering techniques, must be employed (Serban, et al., 2020). The use of software engineering techniques in the development of AI systems is an emerging area of '**IS for AI**', i.e. how to use software engineering techniques to build intelligent systems with adequate quality and productivity. In fact, the '1st Workshop on AI Engineering - Software Engineering for AI' was launched by IEEE and ACM in 2021. The SEI calls this research area 'Artificial Intelligence Engineering[1]' *the result of 'combining*

[a] https://orcid.org/0000-0001-7962-1042
[b] https://orcid.org/0000-0001-7999-9870
[c] https://orcid.org/0000-0003-2155-7409
[d] https://orcid.org/0000-0001-9116-2870
[e] https://orcid.org/0000-0002-7212-8279

[1] https://www.sei.cmu.edu/our-work/artificial-intelligence-engineering/

Oviedo Lama, J., Guerrero Sosa, J. D. T., Rodríguez Monje, M., Romero Chicharo, F. P., Piattini Velthuis and M.
Towards Quality Assessment of AI Systems: A Case Study.
DOI: 10.5220/0013521700003964
In *Proceedings of the 20th International Conference on Software Technologies (ICSOFT 2025)*, pages 254-261
ISBN: 978-989-758-757-3; ISSN: 2184-2833

the principles of systems engineering, software engineering, computer science and human-centred design to create AI systems according to human needs for mission outcomes'.

The use of software engineering techniques for the construction of AI systems is based on the principles established by Horneman (Horneman, et al., 2019), who can be called the father of this new area, considering that AI systems are software-intensive systems, and therefore:

- 'Established principles for designing and deploying quality software systems that meet their mission objectives on time should be applied to engineering AI systems.'
- 'Teams should strive to deliver functionality on time and with quality, design for important architectural requirements (such as security, usability, reliability, performance and scalability), and plan for system maintenance throughout the life of the system.'

While 'Artificial intelligence (AI) is pervasive in today's landscape, there is still a lack of software engineering expertise and best practices in this field' (van Oort, et al., 2021), although developers of AI-intensive systems are aware of the need to employ these techniques, they also recognise that traditional software engineering methods and tools are neither adequate nor sufficient on their own and need to be adapted and extended (Lavazza & Morasca, 2021).

For all these reasons, it is necessary to evolve the methods for quality assurance of 'software 2.0' towards what can be called 'MLware' (Borg, 2021). Although techniques or practices for quality assurance of ML models or ML-based systems are already beginning to emerge (Hamada, et al., 2020), such as the use of white box and black box testing (Gao, et al., 2019), it is necessary to ensure that the tests performed on these AI systems follow established quality standards and criteria: performance, reliability, scalability, security, etc. Therefore, quality assurance of AI-based systems is an area that has not yet been well explored and requires collaboration between the SE and AI research communities and currently, there is a lack of (standardised) approaches for quality assurance of AI-based systems, being essential for their practical use (Felderer & Ramler, 2021).

Therefore, in the new area of Artificial Intelligence Engineering, challenges have arisen related to the definition and guarantee of the behavioural and quality attributes of these systems and applications, which have led us to work on the development of an environment that allows the evaluation of the functional suitability of AI systems. Therefore, the main objective of this article is to present the results obtained in the evaluation of the functional suitability of an artificial intelligence system using the built environment (Oviedo, et al., 2024). In addition, the differences and difficulties encountered in the evaluation of an AI system, compared to the evaluation of a traditional software system, are also presented. For all this, the rest of the article is structured as follows: section 2 presents a summary of the importance of quality in AI systems and the proposals that exist for evaluating the quality of these systems. Section 3 describes the evaluation environment based on ISO/IEC 25000 family standards used to carry out the evaluation. Section 4 presents the AI system evaluated, and the assessment performed. Section 5 presents the conclusions obtained with this work and the future lines of work and research.

2 QUALITY IN AI SYSTEMS

For the quality assurance of artificial intelligence systems, as is done in Software Engineering, it is necessary to consider several dimensions (Borg, 2021), such as the quality of the data, the quality of the AI system itself, the quality of the development processes, etc. However, as mentioned above, existing work on software quality needs to be adapted to the particularities of artificial intelligence. This is why in recent years new standards have emerged that are adapted to the particularities of AI systems to address their quality (Oviedo, et al., 2024) (Piattini, 2024).

At the level of AI system development processes, the new ISO/IEC 5338 standard 'Life cycle processes for AI systems' (ISO/IEC, 2023) has emerged, based on ISO/IEC 12207, which lays the foundations for the relevant processes that organisations should follow to ensure the proper development of AI systems. This standard defines 33 processes separated into four groups, where it adapts the software lifecycle processes to specific aspects of AI. To this end, 23 of the 33 processes have been directly adapted from ISO/IEC 12207 and three new processes have been defined: Knowledge Acquisition Process, AI Data Engineering Process and the Continuous Validation Process (Márquez, et al., 2024).

On the other hand, at the level of AI system product quality, several models, techniques and, in some cases, tools have been proposed to assess and ensure the quality of AI systems. However, these works are based on software engineering standards that have not been adapted to the particularities of AI systems or are focused on evaluating the quality in use and data of the systems (Oviedo, 2024). Therefore, in order to address the evaluation of AI systems, as has happened with other cases (sustainability of software products (Calero, et al., 2021), cloud services (Navas, 2016), e-learning systems (Rahman, 2019), etc.), it has been necessary to adapt the quality model defined in the ISO/IEC 25010 standard for software products to the special characteristics of AI systems, thus giving rise to a new quality model contained in the ISO/IEC 25059 standard (ISO/IEC, 2023) in accordance with the framework of the ISO/IEC 25000 standards.

This new ISO/IEC 25059 standard includes the quality characteristics that any AI system must comply with, based on those proposed by ISO/IEC 25010, but presenting certain changes and adaptations in some of the characteristics and sub-characteristics (Oviedo, et al., 2024).

The proposal made by ISO/IEC 25059 is a starting point that must be completed to be applicable in the industry by defining metrics and thresholds to determine the quality values of each of the sub-characteristics and characteristics defined in the standard. In addition, to address the evaluation of AI systems, it is necessary to have an evaluation process and a technological environment to allow the collection of data necessary for the calculation of the new metrics and to obtain the quality value of each of the characteristics that make up the quality model for the AI system.

3 FUNCTIONAL SUITABILITY ASSESSMENT ENVIRONMENT FOR AI SYSTEMS

In order to address the evaluation of AI systems, it is necessary to adapt existing software quality evaluation environments. On the other hand, as there is no work based on the new ISO/IEC 25059 standard (ISO/IEC, 2023), nor where metrics are defined to know the quality of the AI system itself (Oviedo, et al., 2024), it has been necessary to develop an evaluation environment composed of a quality model and metrics, an evaluation process and a set of measurement and evaluation tools, focusing on one of the characteristics defined in the ISO/IEC 25059 standard, Functional Suitability.

3.1 Quality Model for the Functional Suitability of AI Systems

The first element of the developed environment is the quality model for Functional Suitability (Oviedo, et al., 2024), which is based on the ISO/IEC 25059 standard and complemented with a set of properties and metrics that allow obtaining the level of quality from values of the AI system itself. The definition of properties and metrics has been based on the functional suitability model defined in (Rodriguez, et al., 2016), but its adaptation has been necessary given that Functional Suitability is one of the quality characteristics of the ISO/IEC 25059 standard that present changes with respect to the ISO/IEC 25010 standard (ISO/IEC, 2011). Specifically:

- The sub-characteristic Functional Correctness has been modified given that in AI systems correct results are not produced in the totality of the executions, which has led to the need to establish new metrics for its evaluation, as well as to adapt the measurement procedure.
- The sub-characteristic Functional Adaptability is a new sub-characteristic, so it has been necessary to define new Quality Properties and metrics to address the evaluation of this sub-characteristic.
- At the Characteristic level it has been necessary to make changes in the thresholds and ranges in order to adapt to the changes made in the sub-characteristics.

Figure 1 presents the quality model defined for the functional suitability of AI systems, showing each of the properties that influence each sub-characteristic.

3.2 Evaluation Process

The second element of the environment is the Assessment Process (Oviedo, et al., 2024). The assessment process is based on the ISO/IEC 25040 standard (ISO/IEC, 2023) which already proposes an assessment model, although this process has had to be adapted to address the assessment of the Functional Suitability of IA Systems, in Figure. 2 presents the activities and tasks that make up this process.

3.3 Assessment Tool

The third element is the software environment (Oviedo, et al., 2024), developed to automate the measurement and evaluation of functional suitability in AI systems. The environment has four components:

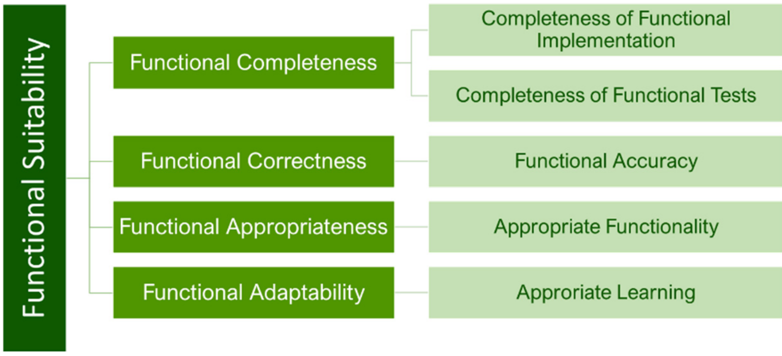

Figure 1: Functional Suitability Quality Model for AI Systems.

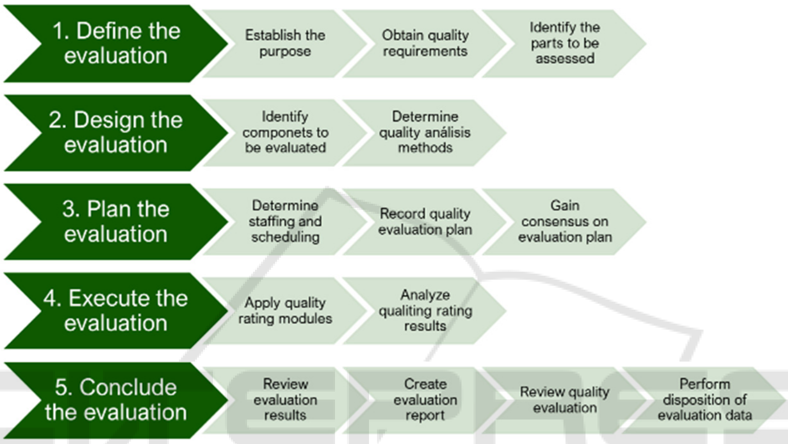

Figure 2: Functional Suitability Assessment Process for AI Systems (Oviedo, et al., 2024).

- Measurement tool: it oversees recording the information of the AI Systems to obtain the necessary metrics for the evaluation of the functional suitability of the AI systems.
- Assessment tool: this tool includes the implementation of the model for assessing functional suitability and a program that performs the necessary calculations to carry out the assessment based on the results of the measurement tool.
- Visualisation tool: this is a visualisation environment where the results are presented.
- Database: where all the information of the evaluations is stored to be used by the rest of the tools of the AI systems evaluation environment.

4 EVALUATION ENVIRONMENT USE CASE

After the development of the evaluation environment, it was necessary to test its applicability on real AI systems to ensure that the metrics, properties, and thresholds defined in the evaluation environment are appropriate. In addition, the use case allowed for the identification of improvements or issues that may exist during the evaluation of an AI system.

To carry out the case study, three AI systems were available as a starting point, whose characteristics are described in Table 1.

Table 1: AI Systems for Case Studies.

Project	Industry	AI System Typology	Accuracy Requirements
Automotive Strike Location	Automotive	fuzzy rules	> 60%.
Skills through videos	-	fuzzy rules	No value
Predict Total Power Consumption for a Customer Type	Energy Distribution	fuzzy rules	> 60%.

In the first place, the 'Skills Assessment through videos' system was discarded for the use case given the large use of computer resources that had to be used to process the video and the subjectivity of the results as it is an assessment of aspects of personality. The system ‘Automotive Strike Location’ was also rejected due to the difficulty of obtaining all the necessary documentation to carry out the evaluation due to the confidentiality of the project.

Finally, the system ‘Prediction of the total power consumption for a type of customer’ was selected as the candidate for the case study given its viability when carrying out the evaluation.

4.1 Presentation of the AI System

The system Prediction of total power consumption for a customer type is a deterministic AI system, based on fuzzy rules. The objective of the system is to predict, from the S02[2] load profiles of the customers of a transformation centre, the level and direction of the effect on power consumption for each of the different types of customers (Public lighting, Family house, Garage, etc.) for a specific day. The predictions are inferred from the fuzzy rules that define the AI model, these rules consider the type of customer and information related to the date such as meteorological and social aspects. The system has two functional requirements, one to determine the level of affectation and the other to determine the direction of affectation. All AI system characteristics of the evaluated product are listed in Table 2.

Table 2: Characteristics of the assessed AI system.

System Type	Fuzzy Rule-based Deterministic		
Nº Fuzzy Rules	266		
Affected Variables	14 variables, including: ▪ Season of the year. ▪ Temperatures: mean, minimum, maximum. ▪ Precipitation. ▪ Hours of sunshine. ▪ Type of day: If it is a specific holiday or the day before or after it, if it is a weekend or if it is a working day or if there was any incident. ▪ Type of customer: Street lighting, Family house, Garage, Business, Sporadic, Church, Industrial building.		
Nº Requirements	2	**Accuracy:**	>60%
Tests Case	2	**Run:**	338 runs of each test case

4.2.1 Defining, Designing and Planning the Evaluation

As this is a use case for validating the functional suitability environment for AI systems, the first three activities of the evaluation process were carried out in parallel and together with the development team of the AI system used for the evaluation.

Initially, the functional suitability model for AI systems was presented to the AI system development team, and each of the measurement elements forming the model (characteristic, sub-characteristic, properties, etc.) were explained in detail. Then, a planning of the activities to be carried out to address the case study was elaborated.

Subsequently, work was undertaken with the development team to identify the elements of the AI system that would be included in the scope of the evaluation and to ensure that these would allow the evaluation of the selected AI system for the use case of the evaluation environment to be addressed. The elements that were identified to carry out the evaluation were: the AI system deployed to be able to be executed, requirements specification, semantic rules, existing documentation of the application and test cases.

In addition, in this initial activity, meetings were also held with the AI system development team to define the dataset needed for the execution of the test cases, since several executions of the same test case must be performed, as established in the evaluation model (Oviedo, et al., 2024). Specifically, to create the dataset for the test cases, it was decided that for each of the 7 types of customers (indicated in Table 2) four dates would be selected at random for each month of the year, with at least one of these dates being on a weekend. This provided a representative dataset that considered different aspects of each of the variables affecting the AI system, as listed in Table 2.

4.2.2 Execute Evaluation

The next activity, the execution of the evaluation, was the most important activity of the use case. It started by recording the requirements specification and the

[2] Hourly load curve incremental values

specification of the test cases (provided by the AI System developer) in the measurement tool of the evaluation environment. The 338 test cases were then run on the system Prediction of total power consumption for one type of customer. Once the two test cases were executed with the different data defined for the tests, the results of these executions were recorded.

Table 3: Metric results in the measurement of Functional Suitability.

METRICS	MEASURE 1	MEASURE 2
Nº of Functional Requirements Specified	2	2
No. of Requirements Implemented	2	2
No. of Requirements Tested	2	2
Coverage	34,61%	41,35%
No. of Exact Functional Requirements	2	2
No. of Improved Requirements	2	2
Specified Usage Targets	1	1
Correct Usage Targets	1	1

After recording the information in the measurement tool, the base metrics used to carry out the evaluation were obtained; these metrics are shown in Table 3. After analysing the results of this first measurement, it was found that most of the metrics obtained adequate results by presenting that the requirements implemented and tested were equal to the number of requirements specified and the degree of accuracy was slightly higher than that established as requirements for the system, which we should remember was 60%. However, most of the fuzzy rules defining the AI model of the evaluated system had not been covered by the test cases that had been executed, despite having performed more than 300 executions of each of the test cases with different data. Therefore, the development team carried out a new review of the AI model rules to define new data and to be able to perform a greater number of runs of the two test cases defined for the AI system in order to validate aspects defined in the rules that had not been taken into account in the first measurement; a total of 338 runs were performed with different data for each of the test cases, thus achieving a slight improvement in this second measurement. Table 3 shows the results of both measurements.

At this point, a meeting was held with the system development team to analyse whether it was possible to obtain new data to cover a greater number of tests, given that after the second measurement only 41.35% of the rules could be tested, given the planning of the evaluation and the deadlines available to the development team, it was decided to evaluate the system with the data from the second measurement and see what results were obtained. The results of the second measurement were processed with the evaluation tool, where the evaluation criteria specified in the model were applied, which made it possible to obtain the quality values of the properties, sub-characteristics and finally the Functional Suitability. Table 4 shows the quality results obtained.

Table 4: Results obtained in the assessment of Functional Suitability.

MEASURING ELEMENT		VALUE
PROPERTIES	Completeness of Functional Implementation	100
	Completeness of Functional Tests	0
	Functional Accuracy	100
	Appropriate Functionality	100
	Appropriate Learning	100
SUB-CARACTERISTICS	Functional Completeness	50
	Functional Correctness	100
	Functional Appropriateness	100
	Functional Adaptability	100
CARACTERISTIC: Functional Suitability		**3**

As can be seen, based on the quality values achieved in the sub-characteristics, a result of 3 out of 5 has been reached for Functional Suitability. Although not an optimal result, it is not a bad result either because 3 out the 4 sub-characteristics have reached the optimum value (100), that the system was developed without considering that it was going to be subjected to a functional suitability evaluation process and the deadlines available to the AI system development team.

4.2.3 Conclude the Evaluation

In the last activity of the evaluation process, the evaluation report was generated, where the results were collected and presented and delivered to the AI system developer. The evaluation results were then reviewed with the AI system developer.

The review of the evaluation results identified the problem with the AI system in relation to Functional Suitability, which was revealed by the quality values obtained in the evaluation of the functional completeness sub-characteristics and their properties. The Functional Completeness sub-characteristic reached an intermediate value due to the result obtained in one of its properties, namely Functional Test Completeness, whose quality level was 0. The reason for this result is because the defined test plan, despite having more than 300 runs with different data, had only managed to cover 41.35% of the semantic rules defined for the AI system. Therefore, the testing process to which the AI system has been subjected does not allow us to ensure that it has been fully tested.

Then, to conclude the evaluation, an analysis of the problem encountered was carried out to identify possible improvements that could be made to the AI system in the future, where it was found that more data was needed to be able to carry out a greater number of test case executions. It was found that the different temperatures that can occur throughout a season were not considered. For example, it was taken into account that in autumn and spring the days would be mild and it has been found that there are rules for these seasons where the temperature is cold or hot.

This improvement was shared with the company that owns the AI system, which found it interesting and committed to carry out an in-depth analysis of the semantic rules to improve its testing process for future developments. However, in this case the company decided not to address in the short term the definition of new datasets to increase the number of executions of the defined test cases mainly for two reasons:

- The users of the AI system were satisfied with the results obtained by the system and therefore did not consider it necessary.
- They did not have enough staff and time to tackle these tasks.

Finally, a review of the results of the case study and all the general experience derived from it was carried out to draw the conclusions presented in the following section.

5 CONCLUSIONS AND FUTURE WORK

The rise of Artificial Intelligence in all facets of everyday life makes it necessary to employ software engineering techniques, such as quality assessment of these systems. While traditional software engineering methods and tools are a starting point for AI systems, they must be adapted to the particularities of these systems. Therefore, this article presents a use case of an evaluation of an AI system using a functional suitability assessment environment, which has been developed considering the particularities of these systems.

Through this case study it has been possible to verify the applicability of the evaluation environment developed to evaluate the functional suitability in AI systems, specifically in deterministic AI systems based on rules. It has also allowed us to verify the permissibility of the evaluation environment, and the need to review and establish more restrictive thresholds that prevent good levels of functional suitability (3 out of 5) from being reached with values of Property to 0.

On the other hand, the use case has allowed us to detect problems related to the use of the evaluation environment. Among the problems detected is the fact that the organisations of these systems do not have a formal specification of the requirements and test cases built to test the AI system, due to the immaturity in the application of software engineering processes in the development of AI systems. Another difficulty encountered in addressing the assessment of functional suitability in AI systems is the large number of executions that must be performed on the defined test cases. This requires a set of input data to ensure that the rules defined for the AI system are covered with a sufficient degree of confidence. Another drawback that has been detected during the use case is related to the different technologies used for the development of AI systems, due to the fact that aspects specific to that technology can make the evaluation more complex. Specifically, in this case, where the system evaluated is a deterministic system based on rules, there have been aspects that have facilitated the evaluation. On the one hand, defining the AI model by means of rules allows us to know the degree of coverage of the tests carried out and, on the other hand, as the results of the system do not change as long as the set of rules of the AI model is not modified, it has facilitated the evaluation of the subcharacteristics of functional adaptability of the system, due to the fact that it has only been necessary to carry out one execution of the set of test cases defined for the system, as there were not going to be any changes in the results. However, these aspects should be considered for future evaluations where the AI systems are not defined by rules, as a greater number of test case executions will be necessary, which could be a possible cause of deviation from the evaluation planning.

With the completion of this study and the conclusions reached, the door is open to carry out other use cases with AI systems with different characteristics, with the aim of adapting and improving the functional suitability assessment environment for AI systems.

ACKNOWLEDGEMENTS

This research has been supported by the ADAGIO project (Alarcos' DAta Governance framework and systems generation) financed by JCCM Consejería de Educación, Cultura y Deportes, and FEDER funds (SBPLY/21/180501/000061), the CASIA project: Calidad de Sistemas de Inteligencia Artificial (EXP. 13/23/IN/002) funded by the Junta de Comunidades de Castilla La Mancha, the AIMM project: Artificial Intelligent Maturity Model (EXP. 13/24/IN/057) funded by the Junta de Comunidades de Castilla La Mancha and the UNION project within the framework of the UCLM Research Plan, 85% co-financed by the European Regional Development Fund (ERDF) UNION (2022-GRIN-34110).

REFERENCES

Borg, M., 2021. The AIQ meta-testbed: pragmatically bridging academic AI testing and industrial Q needs. Vienna, Springer, pp. 66-77.

Calero, C., Moraga, M. & Piattini, M., 2021. Software Sustainability. 1 ed. Cham: Springer.

Comisión Europea, 2020. LIBRO BLANCO sobre la Inteligencia Artificial. Un enfoque europeo orientado a la excelencia y confianza, Bruselas: COM.

Felderer, M. & Ramler, R., 2021. Quality Assurance for AI-Base Systems: Overview and Challeges (Introduccion to Interactive Session).. Viena, Springer, pp. 33-42.

Gao, J., Tao, C., Jie, D. & Lu, S., 2019. What is AI software testing? and Why.. San Francisco, IEEE.

Hamada, K. I. F. M. S. M. M. et al., 2020. Guidelines for Quality Assurance of Machine Learning-based Artificial Intelligence.. Pittsburgh, USA, Word Scientific, pp. 335 -341.

Horneman, A., Melligener, A. & Ozkava, I., 2019. AI Engineering: 11 Foundational Practises. Recommendations for decision makers from expert in software engineering, cibersecurity, and applied artificial intelligence, Pittsburgh: Carnegie Mellon University Software Engineering Institute.

ISO/IEC, 2011. ISO/IEC 25010:2011. Ginebra: International Organization for Standardization.

ISO/IEC, 2023. ISO/IEC 25040:2023. Ginebra: International Organization for Standardization.

ISO/IEC, 2023. ISO/IEC 25059:2023. Ginebra: International Organization for Standardization.

ISO/IEC, 2023. ISO/IEC 5338:2023. Ginebra: International Organization for Standardization.

Lavazza, L. & Morasca, S., 2021. Understanding and Modeling AI-Intensive System Development. Madrid, IEEE, pp. 55-61.

Márquez, R., Rodríguez, M., Romero, F. P. & Piattini, M., 2024. An Artificial Intelligence Maturity Assessment Framework Based on Interantional Standard. Enviado a Engineering Applications of Artificial Intelligence.

Maslej, N. et al., 2024. Intelligence Index Report 2024, Stanford: Stanford University.

Navas, R., 2016. Modelos de Calidad para servicios en el Cloud. Valencia: Universidad Politécnica de Valencia.

Oviedo, J., 2024. CSIA: Evaluación de la Adecuación Funcional de Sistemas de Inteligencia Artificial, Ciudad Real: Grupo Alarcos UCLM.

Oviedo, J., Rodríguez, M. & Piattini, M. 2024. An Environment for the Assessment of the Functional Suitability of AI Systems. Pisa, Springer, pp. 21-34.

Oviedo, J., Rodríguez, M. & Piattini, M., 2024. Entorno para la Evaluación de la Adecuación Funcional en Sistemas IA. Curitiba (Brasil), Curran Associates, pp. 364-371.

Oviedo, J. y otros, 2024. ISO/IEC quality standards for AI engineering. Computer Science Review, 54(1), p. 100681.

Piattini, M., 2024. Gobierno, Gestión y Calidad de la Inteligencia Artificial. 1 ed. Puertollano (Ciudad Real): Grupo Alarcos.

Rahman, A., 2019. Quality Consideration for e-Learing System based on ISO/IEC 25000 Quality Standards. Putrajaya (Malasia), DBLP Computer Science Bibliography, pp. 12-17.

Rodríguez, M., Oviedo, J. & Piattini, M., 2016. Evaluation of Software Product Functional Suitability. A Case Study. Software Quality Professional, 18(3), pp. 18-29.

Serban, A., Koen, v. d. B., Holger, H. H. & Visser, J., 2020. Adoption and Effects of Software Engineering Best Practise in Machine Learning. Bari (Italia), Computer Science, Engineering.

van Oort, B., Cruz, L., Aniche, M. & van Deursen, A., 2021. The Prevalence of Code Smells in Machine Learning projects. Madrid, IEEE, pp. 35-42.