Towards an Intelligent Platform for Supporting Quality Assessment in Blockchain Systems

Raphael Saraiva
State University of Ceara
Fortaleza, Ceará, Brazil
raphael.saraiva@aluno.uece.br

Allysson Allex Araújo
Federal University of Cariri
Juazeiro do Norte, Ceará, Brazil
allysson.araujo@ufca.edu.br

Ronnie de Souza Santos
University of Calgary
Calgary, AB, Canada
ronnie.desouzasantos@ucalgary.ca

Pamella Soares
State University of Ceara
Fortaleza, Ceará, Brazil
pamella.soares@aluno.uece.br

Jerffeson Souza
State University of Ceara
Fortaleza, Ceará, Brazil
jerffeson.souza@uece.br

Abstract

The emergence of blockchain has transformed multiple sectors by enabling security, transparency, and decentralization. Yet, evaluating the quality of blockchain-based systems remains a complex task due to their distributed nature and the absence of specific standards. This study proposes a research design for the development and evaluation of an intelligent platform that supports quality assessment in blockchain systems. The solution integrates a curated catalog of software quality metrics aligned with ISO/IEC 25010 and a chatbot powered by Retrieval-Augmented Generation (RAG) that assists researchers and practitioners in identifying and applying appropriate metrics. Grounded in the Design Science Research (DSR) methodology, the study combines technical and social analyses to assess aspects such as throughput, latency, usability, and perceived usefulness. The study advances the field by introducing an intelligent platform and an evaluation framework designed to promote standardization, transparency, and shared understanding in blockchain quality assessment.

CCS Concepts

• **General and reference → Metrics**; • **Information systems → Data management systems**.

Keywords

Blockchain; Software Quality; ISO/IEC 25010; RAG; Metrics Catalog.

ACM Reference Format:
Raphael Saraiva, Allysson Allex Araújo, Ronnie de Souza Santos, Pamella Soares, and Jerffeson Souza. 2026. Towards an Intelligent Platform for Supporting Quality Assessment in Blockchain Systems. In *8th Workshop on Emerging Trends in Software Engineering for Blockchain (WETSEB '26), April 12–18, 2026, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3786157.3788566

ACM ISBN 979-8-4007-2389-6/2026/04
https://doi.org/10.1145/3786157.3788566

1 Introduction

In recent years, blockchain technology has emerged as a promising innovation, potentially contributing to overcoming various challenges and being compared to the early impact of the Internet and the Web in their initial days [26]. Blockchain consists of a distributed ledger, first introduced by Nakamoto [19] as part of the system proposed to enable the operation of Bitcoin, a decentralized digital cryptocurrency that records financial transactions without the need for intermediaries, ensuring security and transparency through decentralized consensus.

According to Beck et al. [3], blockchain is a highly transparent, secure, and resilient record of digital events. Over time, the concept of blockchain has evolved beyond cryptocurrencies, driven by the emergence of smart contracts, introduced by Buterin et al. [5] as self-executing code that enables digital agreements to operate without intermediaries. This advancement has established blockchain as a general-purpose technology with applications across various sectors. Blockchain has played a central role in transformation, particularly in finance, healthcare, logistics, and governance, where it promotes greater efficiency and reduces operational costs [14]. Its ability to eliminate intermediaries and create a trusted environment for peer-to-peer (P2P) transactions stands out as a distinctive advantage in an increasingly digital world.

Failures in blockchain implementations have undermined this trust, revealing persistent challenges in ensuring robustness and security. In 2022, the Ronin network, used by Axie Infinity, was targeted by an attack that exploited vulnerabilities in validator nodes, resulting in the theft of approximately USD 625 million in cryptocurrencies [7]. Another illustrative case in 2023 involved Atomic Wallet, a decentralized cryptocurrency wallet, which suffered an attack, causing losses exceeding USD 100 million and highlighting the weaknesses in digital asset storage solutions [20]. These incidents demonstrate that, although blockchain holds great potential, its implementation requires continuous investment in quality assessment and robustness to protect users and strengthen the ecosystem.

According to Zhu et al. [37], software quality is directly associated with its ability to adapt to the dynamic demands of the technological environment, encompassing attributes such as maintainability, adaptability, and updatability, which enable the software to evolve continuously to meet new requirements and operational contexts. These aspects are critical to ensuring that blockchain-based systems, in particular, can adjust to constant changes in security, performance, and compliance requirements, minimizing risks and promoting ecosystem reliability.

However, despite the growing set of tools and frameworks for blockchain experimentation and performance evaluation, most existing solutions focus primarily on benchmarking consensus mechanisms, throughput, and scalability [2]. While these tools have advanced the understanding of blockchain efficiency, they rarely address software quality metrics in a broader sense, as defined by international standards such as ISO/IEC 25010 [1]. Moreover, existing blockchain quality frameworks provide mostly static metric descriptions [30], while recent RAG-based software engineering assistants lack an explicit quality model [17]. Consequently, there is still a lack of comprehensive environments that support the systematic assessment of quality attributes (e.g., usability, reliability, etc). These limitations reinforce the need for effective methodologies to support the application of metrics and assist in the quality assessment of blockchain-based systems.

Given this context, selecting appropriate metrics to assess the quality of blockchain systems remains challenging. Traditional software quality metrics require adaptation, as blockchain systems differ from conventional software due to immutability, decentralized execution, consensus protocols, probabilistic finality, transaction fees, and domain-specific security threats [26]. In addition, quality metrics vary across blockchain platforms, and the lack of a consolidated reference framework complicates their interpretation and application in practice. Moreover, quality assessment in blockchain ecosystems also involves social and collaborative aspects, as the effective use of metrics depends on how researchers and practitioners interpret and apply them in practice.

Building upon research gap, this extended abstract presents the methodological design of a support platform to assist both academia and industry in the evaluation and application of software quality metrics in blockchain-based systems. The research design implements an approach that provides structured guidance for understanding, selecting, and applying quality metrics within decentralized environments of high complexity and heterogeneity. The platform integrates two components: a metrics catalog aligned with the ISO/IEC 25010 standard, gathering structured information on quality attributes, formulas, application criteria, and relevant elements; and a chatbot to assist users in identifying, organizing, and applying these metrics in blockchain contexts. The chatbot employs Language Models (LLMs) combined with Retrieval-Augmented Generation (RAG), enabling adjustment of responses based on information from the catalog. The platform provides interactive and contextualized support, facilitating access to application of metrics that suit the needs of each scenario.

From a scientific standpoint, this paper adopts a Design Science Research (DSR) approach that encompasses both the technical evaluation of the platform and the analysis of user interaction and perception during its use. This approach allows examining both the platform's technical performance and the human factors that influence its adoption, its scientific and practical value. The proposed artifact may help overcome barriers to the use of quality metrics and supports more consistent evaluation of blockchain systems.

The contributions of this work are presented in three parts. (i) a methodological approach for platforms that support quality assessment in blockchain; (ii) a proposed integrated artifact, consisting of a metrics catalog and a RAG-based chatbot; (iii) a social and technical evaluation plan organized into stages and instruments, offering a replicable basis for assessing blockchain system quality with transparency and reproducibility. These contributions address the challenges of evaluating decentralized software and align with the emphasis on improving approaches and engineering practices for blockchain systems.

2 Related Work

Recently, researchers have started using question-and-answer systems powered by artificial intelligence to make information about blockchain easier to access [4, 17]. These systems use LLMs that search through stored texts and, when needed, consult external sources on the web to improve the accuracy and reliability of their answers. For example, Mansurova et al. [17] developed a blockchain-focused chatbot that integrates document segmentation, semantic search, and conversational memory to improve access to technical information and support user learning in this domain. In more practical applications, similar systems are already being used in blockchain-based traceability platforms, where users can ask questions in natural language about the origin and quality of products throughout the production chain [4].

Security has been a primary focus of research on blockchain quality. Recent work has explored chatbots and LLMs to support smart contract auditing, compliance, and vulnerability detection, which are directly related to security and reliability [6, 16, 35]. Complementary studies apply machine learning techniques to the qualitative analysis of smart contract vulnerabilities, highlighting potential and limitations in security assessment [28]. Another line of research explores chatbots to support the specification and understanding of smart contracts, aiming to reduce the gap between natural-language requirements and smart contract code [27].

Differently from previous works, this research introduces an application of LLM+RAG techniques to support the quality assessment of blockchain-based systems. The study includes the design of a catalog of metrics aligned with the ISO/IEC 25010 standard, integrated into a chatbot interface that assists users in identifying and interpreting indicators. Rather than using RAG to provide domain-specific explanations or access to unstructured documentation, the chatbot is coupled with this metrics catalog and operates as a decision-support mechanism for selecting, contextualizing, and applying software quality metrics. The approach advances software quality analysis in blockchain contexts by incorporating dimensions such as performance, reliability, and usability, complementing the emphasis on security. It employs an evaluation approach that combines quantitative performance analysis focused on retrieval accuracy, contextual relevance, and response stability with qualitative methods such as participant observation, usability assessment, and perception analysis following the Technology Acceptance Model (TAM) [9] to examine how technical and human factors influence system quality.

3 Methodological Procedures

This section presents the nature of the research and the methodological procedures adopted in each stage, organized to demonstrate the research objectives, the artifact design, and the respective evaluation activities.

3.1 Research Characterization

This research was conducted under the Design Science Research (DSR) approach, which focuses on the creation and evaluation of artifacts to address real-world problems in the field of information technology [12]. DSR is widely used to develop solutions that are applicable to both academia and industry. In this study, the methodology was employed to design and validate a platform aimed at supporting the quality assessment of blockchain systems.

Figure 1 illustrates the overall structure of this research, detailing its main phases along with their corresponding processes and methods. Based on the DSR protocol suggested by Vaishnavi and Kuechler [33], our methodological procedure employed follows five main phases within the context of DSR. The following sections present each phase, describing the objectives and methodological procedures adopted.

	PROCESS	METHOD
PHASE 1 PROBLEM AWARENESS	Identifying the state of the art of blockchain metrics	Systematic Mapping Study (SMS)
PHASE 2 SUGGESTION	Creationg of the catalog for blockchain metrics	Creation of the catalog based on the data collected through SMS
PHASE 3 DEVELOPMENT	Tool development with catalog metrics	Implementation of front-end, back-end, and integration with LLM models
PHASE 4 EVALUATION	Quantitative and Qualitative experimental evaluation	QUANTITATIVE EVALUATION performance analysis QUALITATIVE EVALUATION interviews & usability testing
PHASE 5 CONCLUSION	Reflection on general knowledge based on experimental evaluations	

Figure 1: Phases of the Design Science Research Process.

Initially in Phase 1 (**Problem Awareness**), a Systematic Mapping Study (SMS) was conducted following established guidelines [13, 25] to identify and classify software quality metrics used in blockchain-based systems. The SMS analyzed 128 primary studies retrieved from Scopus, IEEE Xplore, ScienceDirect, and ACM Digital Library, identifying 160 distinct metrics reported in empirical evaluations and classifying them according to ISO/IEC 25010 quality characteristics and evaluation levels.The complete results of this mapping study were published in [30].

The Suggestion Phase centered on the creation of the metrics catalog, conceived as a core component of the proposed platform. This phase was based on the results obtained from the SMS [30], which made it possible to identify gaps and best practices related to quality assessment in blockchain systems. The structure of the catalog was inspired by the approach proposed by Olsina et al. [22], whose systematic organization of attributes and association with quality models provided guidance for defining the catalog's structure and specifications. This methodology enabled the organization and formalization of the catalog using the data and evidence collected in Phase 1, ensuring consistency and alignment with established references. Concrete examples of fully documented metrics, including their definitions, objectives, formulas, scales, and usage protocols, are available in our supplementary repository [31].

In the **Development Phase**, the platform was implemented by consolidating the technical and conceptual foundations established in the previous stages. The development included the creation of an API using AdonisJS [34], responsible for managing communication between components and handling data processing, as well as a web interface built with ReactJS [18]. Integration with large language models was performed directly at the API level, which communicates with locally hosted models via Ollama [32] and with the ChatGPT API, enabling the adoption of the Retrieval-Augmented Generation (RAG) approach.

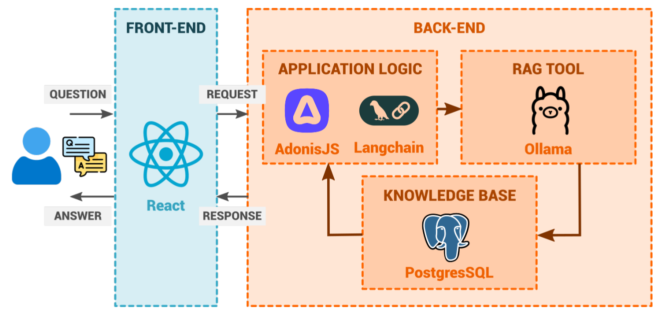

Figure 2: Architectural model of the platform.

As illustrated in Figure 2, the chatbot processes user queries by retrieving relevant information from the PostgreSQL knowledge base via the RAG component and combining it with the language model to generate contextualized responses in real time.

In the **Evaluation Phase**, an empirical analysis will be conducted to examine both technical and social aspects of the platform. The *quantitative evaluation* will focus on performance analysis, assessing retrieval accuracy, contextual relevance, and response stability to verify the platform's effectiveness, efficiency, and reliability. The *qualitative evaluation*, in turn, will rely on interviews and usability testing to analyze user perception and acceptance of the platform based on the Technology Acceptance Model (TAM) [9]. The model considers four constructs: perceived usefulness, perceived ease of use, attitude toward use, and behavioral intention, which explain how individuals decide to adopt and continue using a technology. The evaluation design will be detailed in the following subsection, as this phase remains to be conducted and represents the main component of the study.

Finally, in the **Conclusion Phase**, the study will consolidate the evidence obtained throughout the research and reflect on its theoretical and practical implications. This stage aims to articulate how the findings contribute to advancing the understanding of quality assessment in blockchain systems and to discuss the broader applicability of the proposed approach within both academic and practical contexts.

3.2 Evaluation Design

The evaluation follows a mixed-methods framework [11], combining quantitative system results with qualitative evidence from

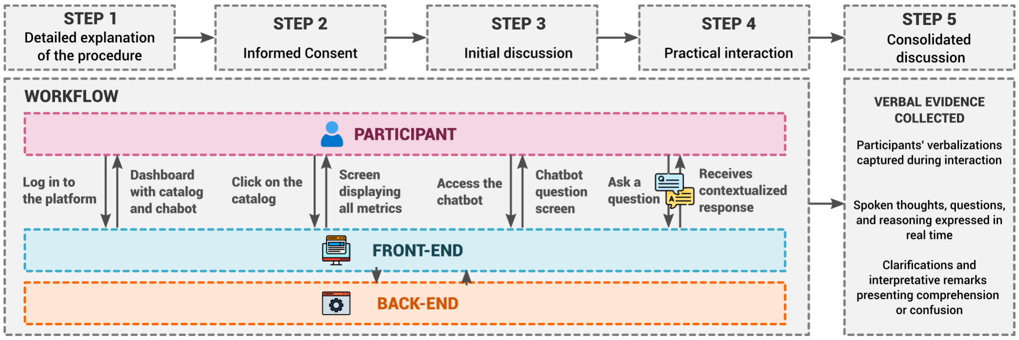

Figure 3: Workflow of the Qualitative Evaluation Phase.

participant observations and questionnaires to provide a comprehensive view of platform effectiveness. The following subsections describe the quantitative and qualitative evaluations in detail.

3.2.1 Quantitative Evaluation. This evaluation assesses the tool's ability to produce accurate and contextually faithful responses using standard metrics for language models and RAG systems, including BLEU [24], ROUGE [15], BERTScore [36], and RAGAS [10]. BLEU and ROUGE focus on linguistic and structural aspects such as n-gram correspondence and lexical coverage, while BERTScore captures semantic similarity between generated responses and reference texts [36]. RAGAS evaluates the contextual quality of RAG systems across *faithfulness*, *context precision*, *context recall*, and *answer relevance* [10], assessing both fluency and faithfulness to retrieved evidence.

We will conduct the evaluation using the language models Llama2 and Llama3 [32]. Their open-weight nature and shared architecture will enable a controlled and reproducible comparison of RAG behavior across model generations, consistent with model-agnostic RAG evaluation frameworks such as RAGAS [10]. In addition, GPT-5 [23] will be included as a representative proprietary large-scale model serving as an external performance baseline. Statistical significance will be assessed using the Wilcoxon (WC) test and effect size will be quantified with the Vargha-Delaney (A_{12}) statistic.

3.2.2 Qualitative Evaluation. Interviews & Usability Testing will be adopted for this evaluation, following the methodological principles proposed by Nielsen [21]. Each participant, within a simulated scenario, will perform predefined tasks by interacting with the artifact's interface. This methodology will stand out for enabling a personalized analysis of how the solution is used in different contexts. It is a testing protocol commonly applied in usability studies, allowing the collection of direct and detailed feedback from each participant, as discussed by Rubin and Chisnell [29]. The purpose of this evaluation is to capture multiple perspectives on participants' interaction with the tool, providing an understanding of aspects as usability, clarity of metrics, and applicability of the solution. Individual interviews were chosen as the primary method, allowing for a detailed exploration of users' experiences and perceptions regarding the proposed solution.

We defined the participation criteria based on profiles representing the target audience of the tool, including individuals aged 18 or older with technical background and prior experience in blockchain development and software quality, and at least one year of professional or research experience in these areas. Participants will be recruited through professional networks and LinkedIn, and selected candidates will receive individual email invitations describing the study, its stages, expected duration, and potential risks and benefits.

Figure 3 presents the structure of the qualitative evaluation process, organized into five main stages. This phase follows an empirical protocol inspired by usability testing practices [21], enabling systematic assessment of user interaction and perception under realistic conditions. In **Step 1 (Briefing)**, the researchers will explain the experimental procedures in detail to ensure that participants clearly understand the research objectives and the activities they would perform, followed by **Step 2 (Consent)**, in which participants will read and sign the Informed Consent Form (ICF), confirming their agreement to participate in the study and their understanding of all ethical aspects involved.

In **Step 3 (Discussion)** will involve an initial discussion in which each participant was invited to share their perceptions, prior knowledge, and practical experiences related to the use of metrics in the blockchain context. The purpose of this stage was to introduce the topic and encourage participants to express their opinions freely, identifying perceived challenges and opportunities independently.

During **Step 4 (Interaction)**, participants will engage directly with the platform, completing evaluation activities that simulate real quality-assessment scenarios. They will explore the metric catalog, consult the chatbot to obtain metric recommendations, and interpret the generated responses to solve predefined tasks. Throughout this stage, the researcher will monitor user interaction to capture evidence related to usability, clarity of metrics, and applicability of the solution. The process will be guided by the participant observation technique [8] and the Think-Aloud method [29], which will encourage participants to verbalize their thoughts, questions, and reasoning during task execution. The verbal evidence collected in these sessions, illustrated in Figure 3, will reveal how participants interpret metric recommendations and navigate between the catalog and chatbot, exposing moments of understanding, confusion, and perceived usefulness. This analysis will complement the structured assessment to be conducted in Step 5.

Finally, in **Step 5 (Questionnaire)**, three complementary sections will be administered to consolidate participants' perceptions after using the platform. The first is a *demographic questionnaire*, used to describe participants' profiles. The second is the *evaluation questionnaire*, structured according to the four constructs of TAM previously introduced in Section 3.1: perceived usefulness, perceived ease of use, attitude toward use, and behavioral intention. Each section includes Likert-scale statements adapted from traditional TAM instruments to fit the specific context of the proposed platform, such as "*The chatbot provided relevant answers to my questions about the metrics*". This item is a contextualized adaptation that measures perceived usefulness in terms of the adequacy and relevance of the chatbot's responses. The third section is an *open feedback form*, allowing participants to provide additional comments or suggestions for improving the platform. All preliminary materials, including the questionnaires and consent form, are available in our supplementary repository [31].

4 Final Remarks

We present a research approach for the development and evaluation of a platform that integrates an ISO/IEC 25010-based metrics catalog and a RAG-enabled chatbot, aimed at supporting quality assessment in blockchain systems. Preliminary results indicate that the RAG-based approach achieved competitive BLEU, ROUGE and BERTScore values while producing responses with richer and more informative content grounded in the knowledge base.

The next phase involves conducting quantitative and qualitative studies to empirically assess the effectiveness of the proposed approach. As future work, we also plan to incorporate platform-aware metric interpretation in the RAG component, since different blockchain architectures such as Ethereum and Solana impose distinct performance, consensus, and security assumptions that can affect how quality metrics should be interpreted.

References

[1] 2023. ISO/IEC 25010:2023 — Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models. https://www.iso.org/standard/78176.html Accessed: 2024-11-12.

[2] Adel Albshri, Ali Alzubaidi, Bakri Awaji, and Ellis Solaiman. 2022. Blockchain simulators: A systematic mapping study. In *2022 IEEE International Conference on Services Computing (SCC)*. IEEE, 284–294.

[3] Roman Beck, Michel Avital, Matti Rossi, and Jason Bennett Thatcher. 2017. Blockchain technology in business and information systems research. 381–384 pages.

[4] José Benzinho, João Ferreira, Joel Batista, Leandro Pereira, Marisa Maximiano, Vítor Távora, Ricardo Gomes, and Orlando Remédios. 2024. LLM Based Chatbot for Farm-to-Fork Blockchain Traceability Platform. *Applied Sciences* 14, 19 (2024), 8856. doi:10.3390/app14198856

[5] Vitalik Buterin et al. 2014. A next-generation smart contract and decentralized application platform. *white paper* 3, 37 (2014), 2–1.

[6] Chong Chen, Jianzhong Su, Jiachi Chen, Yanlin Wang, Tingting Bi, Jianxing Yu, Yanli Wang, Xingwei Lin, Ting Chen, and Zibin Zheng. 2023. When ChatGPT Meets Smart Contract Vulnerability Detection: How Far Are We? *arXiv preprint arXiv:2309.05520* (2023). arXiv:2309.05520 [cs.SE] doi:10.48550/arXiv.2309.05520

[7] conjur. 2023. Ataques a blockchain e a responsabilidade civil. https://www.conjur.com.br/2022-jun-02/opiniao-ataques-blockchain-responsabilidade-civil/ Accessed: 2024-11-12.

[8] Janet Cooper, Rachael Lewis, and Christine Urquhart. 2004. Using participant or non-participant observation to explain information behaviour. *Information Research* (2004), 9–4.

[9] Fred D Davis, Richard P Bagozzi, and Paul R Warshaw. 1989. Technology acceptance model. *J Manag Sci* 35, 8 (1989), 982–1003.

[10] Shahul Es, Jithin James, Luis Espinosa Anke, and Steven Schockaert. 2024. Ragas: Automated evaluation of retrieval augmented generation. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*. 150–158.

[11] Jennifer C Greene, Valerie J Caracelli, and Wendy F Graham. 1989. Toward a conceptual framework for mixed-method evaluation designs. *Educational evaluation and policy analysis* 11, 3 (1989), 255–274.

[12] Alan R Hevner, Salvatore T March, Jinsoo Park, and Sudha Ram. 2004. Design science in information systems research. *MIS quarterly* (2004), 75–105.

[13] Staffs Keele et al. 2007. *Guidelines for performing systematic literature reviews in software engineering*. Technical Report. Technical report, ver. 2.3 ebse technical report. ebse.

[14] Mahtab Kouhizadeh, Sara Saberi, and Joseph Sarkis. 2021. Blockchain technology and the sustainable supply chain: Theoretically exploring adoption barriers. *International journal of production economics* 231 (2021), 107831.

[15] Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*. 74–81.

[16] Wei Ma, Daoyuan Wu, Yuqiang Sun, Tianwen Wang, Shangqing Liu, Jian Zhang, Yue Xue, and Yang Liu. 2024. Combining fine-tuning and llm-based agents for intuitive smart contract auditing with justifications. *arXiv preprint arXiv:2403.16073* (2024).

[17] Aigerim Mansurova, Aliya Nugumanova, and Zhansaya Makhambetova. 2023. Development of a Question Answering Chatbot for Blockchain Domain. *Scientific Journal of Astana IT University* 15 (Sept. 2023), 27–39. doi:10.37943/15XNDZ6667

[18] Meta and Contributors. 2025. React - A JavaScript library for building user interfaces. https://react.dev Accessed: January 18, 2025.

[19] Satoshi Nakamoto. 2008. Bitcoin: A peer-to-peer electronic cash system. *Satoshi Nakamoto* (2008).

[20] Blockchain News. 2023. Atomic Wallet Hack: Over 100 Million Stolen. https://www.blockchainnews.com/atomic-wallet-hack/ Acessado em 8 de dezembro de 2024.

[21] Jakob Nielsen. 1994. *Usability engineering*. Morgan Kaufmann.

[22] Luis Olsina, Guillermo Javier Lafuente, and Oscar Pastor. 2002. Designing a Catalogue for Metrics. In *Proceedings of the International Conference on Web Engineering (ICWE)*. Springer, 348–358. https://www.researchgate.net/publication/338677862_Designing_a_Catalogue_for_Metrics

[23] OpenAI. 2025. GPT-5 Technical Overview. https://www.openai.com. Accessed: 2025-01-19.

[24] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*. 311–318.

[25] Kai Petersen, Robert Feldt, Shahid Mujtaba, and Michael Mattsson. 2008. Systematic mapping studies in software engineering. In *12th international conference on evaluation and assessment in software engineering (EASE)*. BCS Learning & Development.

[26] Simone Porru, Andrea Pinna, Michele Marchesi, and Roberto Tonelli. 2017. Blockchain-oriented software engineering: challenges and new directions. In *2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C)*. IEEE, 169–171.

[27] Ilham A. Qasse, Shailesh Mishra, and Mohammad Hamdaqa. 2021. iContractBot: A Chatbot for Smart Contracts' Specification and Code Generation. In *2021 IEEE/ACM 3rd International Workshop on Bots in Software Engineering (BotSE)*. IEEE, 35–38. doi:10.1109/BotSE52550.2021.00015

[28] Dalila Ressi, Alvise Spanò, Lorenzo Benetollo, Michele Bugliesi, Carla Piazza, and Sabina Rossi. 2025. Vulnerability Detection in Solidity Smart Contracts via Machine Learning: A Qualitative Analysis. *Blockchain: Research and Applications* (2025), 100390.

[29] Jeffrey Rubin and Dana Chisnell. 2011. *Handbook of usability testing: How to plan, design, and conduct effective tests*. John Wiley & Sons.

[30] Raphael Saraiva, Allysson Allex Araújo, Pamella Soares, João Carlos Pontes, and Jerffeson Souza. 2025. Metrics for Quality Assessment in Blockchain-based Systems: A Systematic Mapping Study. *Simpósio Brasileiro de Sistemas de Informação (SBSI)* (2025), 545–554.

[31] R. Saraiva, A. A. Araújo, Santos S. R., P. Soares, and J. Souza. 2025. Repository of "Supplementary materials for Towards an Intelligent Platform for Supporting Quality Assessment in Blockchain Systems". https://doi.org/10.5281/zenodo.17536749.

[32] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).

[33] Vijay Vaishnavi and William Kuechler. 2004. Design research in information systems. (2004).

[34] Harminder Virk and Contributors. 2025. AdonisJS - A Fully Featured Web Framework for Node.js. https://adonisjs.com Accessed: January 18, 2025.

[35] Zhiyuan Wei, Jing Sun, Zijiang Zhang, Xianhao Zhang, Meng Li, and Zhe Hou. 2024. LLM-SmartAudit: Advanced Smart Contract Vulnerability Detection. *arXiv preprint arXiv:2410.09381* (2024). arXiv:2410.09381 [cs.CR] doi:10.48550/arXiv.2410.09381

[36] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675* (2019).

[37] Xin Zhu, Lei Wang, and Jian Yang. 2022. Software Quality and Maintenance: Challenges and Strategies. *Journal of Software Engineering* 30, 4 (2022), 123–135.