Universal Library of Engineering Technology | Research Article

ISSN: 3064-996X | Volume 2, Issue 4

Open Access | PP: 57-62

DOI: https://doi.org/10.70315/uloap.ulete.2025.0204010

Application of DORA Metrics for Assessing the Effectiveness of Continuous Integration and Delivery (CI/CD) Processes

Oleksiy Davydenko

Delivery Director, Trinetix.

Abstract

The article is devoted to the analysis of the problems of measuring the performance of continuous integration and delivery (CI/CD) processes under conditions of scaling engineering teams and increasing complexity of software and technical ecosystems. The scientific and practical significance of the study is determined by the transition from linear, predominantly quantitative indicators to multidimensional evaluation systems that establish stable links between technical metrics, operational characteristics, and goal setting at the level of business strategy. The paper examines the practice of relying exclusively on DORA metrics; clarifies the boundaries of their correct application in high load, distributed organizational contexts; and explores possibilities for enriching their content by integrating them with the SPACE framework and Site Reliability Engineering (SRE) approaches. Particular attention is paid to the Lean Portfolio Management (LPM) methodology, interpreted as a mechanism for strategic alignment of value streams that makes it possible to directly couple decisions in the area of CI/CD with processes of managing the portfolio of products and initiatives. The aim of the study is to develop an integrated model for evaluating the effectiveness of CI/CD processes that ensures the alignment of the speed of delivering changes with the requirements for the resilience and reliability of the systems being developed and operated. To achieve this aim, methods of comparative and systems analysis are applied. In the concluding part, the concept of two loop measurement is formulated and substantiated as a model in which managerial decisions are based on the interpretation of correlation relationships between metrics of engineering practice performance and indicators of team satisfaction and resilience. The proposed approach can be used as a methodological and practical guideline for chief technology officers (CTOs), heads of development units, and SRE specialists responsible for the design and development of systems for measuring the effectiveness of engineering activities in scalable organizations.

Keywords: *DORA Metrics, CI/CD Efficiency, SPACE Framework, Site Reliability Engineering, Error Budget, Lean Portfolio Management.*

INTRODUCTION

The relevance of the stated topic is determined not only by the growing architectural and technological complexity of modern software systems, but also by the intensifying institutional pressure to shorten the time-to-market cycle for digital products while simultaneously maintaining an acceptable level of reliability and controllable risk exposure. As engineering units scale up (including growth from about 40 to 120 and more specialists), traditional performance indicators lose their sensitivity to the transformational changes taking place in the organization and cease to reflect the causal relationship between the activities of development teams and the achieved business outcomes [2]. An exclusive focus on quantitative characteristics of development speed, while ignoring process stability and the human factor, leads to the accumulation of technical debt, a decline in maintenance quality, and an increased likelihood of employee burnout.

Under these conditions, there arises a need for a holistic system of metrics capable of linking the operational characteristics of engineering practices with the strategic objectives of the organization, including risk management, regulatory compliance, and the balance of interests of key stakeholders. The limitations of fragmented approaches to assessing CI/CD performance manifest themselves, in particular, in the fact that operational indicators often exist in isolation from corporate governance circuits, are not incorporated into managerial decision-making processes, and are in practice not used as a basis for adjusting product and portfolio strategies [1, 9].

The aim of this study is to substantiate and design a

Citation: Oleksiy Davydenko, "Application of DORA Metrics for Assessing the Effectiveness of Continuous Integration and Delivery (CI/CD) Processes", Universal Library of Engineering Technology, 2025; 2(4): 57-62. DOI: https://doi.org/10.70315/uloap.ulete.2025.0204010.

comprehensive approach to assessing the effectiveness of continuous integration and delivery (CI/CD) practices, which integrates DORA operational metrics, qualitative indicators structured within the SPACE framework, as well as instruments of strategic portfolio management (LPM).

To achieve this aim, the following **objectives** are formulated:

– to analyze the applicability of DORA metrics and the SPACE conceptual model for assessing the maturity of software delivery processes in large organizations implementing large-scale digital transformation programs [1, 9];

– to investigate the role of error-handling policies and service level objectives (SLOs) as legal and managerial instruments for balancing the acceleration of the development cycle with the maintenance of stable operation of information systems [6, 10];

– to formulate principles and mechanisms for embedding delivery metrics into strategic portfolio management (LPM) procedures in line with the SAFe methodology, including escalation circuits, resource reallocation, and initiative prioritization [7].

The scientific novelty of the study lies in the proposal of a hybrid monitoring model in which DORA metrics are interpreted not only as indicators of operational performance but also as formalized triggers for initiating managerial interventions at the portfolio management level.

The author's hypothesis is that high effectiveness of software delivery is fundamentally attainable only when a dual-loop feedback system is in place, in which indicators of change velocity (DORA) are systematically counterbalanced by metrics of team well-being and sustainability (SPACE), as well as by reliability constraints set through Error Budgets and the corresponding SLOs. Such a configuration makes it possible to interpret delivery speed as a derived characteristic of the stability of the sociotechnical system rather than as an independent goal, which is critically important for the long-term development of digital platforms.

MATERIALS AND METHODS

The methodological basis of the study was formed on the foundation of a comprehensive interdisciplinary approach combining tools of a systematic review of the scientific literature. The theoretical and empirical basis of the study consisted of the results of a targeted search for scientific publications and industry reports in leading international abstract and citation databases. The temporal boundaries of the sample were set in the range 2021–2025, which made it possible to focus on the most relevant concepts and methods for measuring the effectiveness of software delivery and the management of engineering organizations. Search strategies were constructed using combined logical operators and included the following key phrases: DORA metrics, Software Delivery Performance, SPACE framework, Lean Portfolio Management, Site Reliability Engineering, DevOps efficiency, Error Budget policy. To reduce the risk of systematic bias, a principle was applied that presupposes the analysis of reference lists in relevant works and the sequential expansion of the corpus of sources by incorporating the studies cited in them.

Additionally, when interpreting the results, the method of methodological triangulation was applied, involving the comparison of conclusions obtained on the basis of academic research with the provisions of international standards and practice-oriented reports of industry leaders. This made it possible not only to identify the dominant trends in the development of metric-based management in Agile/DevOps environments, but also to delineate the limits of applicability of individual groups of metrics in large corporate structures with distributed legal and managerial competencies, including issues of responsibility for software quality, compliance with SLA and the fulfillment of contractual obligations in the field of digital services.

RESULTS

In the paradigm of evaluating software delivery effectiveness, the emphasis shifts from metrics of the amount of work performed (number of lines of code, person-hours, and similar measures) to indicators describing the business value being created and the predictability of the release cycle. In this context, the most common standard is the set of four DORA metrics: deployment frequency, the lead time for changes from commit to production, the percentage of releases resulting in incidents, and the time to restore service after failures. Empirical studies demonstrate a stable positive correlation between high values of these indicators and the dynamics of key business metrics and the market capitalization of technology companies [2, 11].

The SPACE framework proposes considering five interrelated dimensions: satisfaction, actual performance, observable activity, quality of communication, and workflow efficiency [1]. This approach has undergone validation, including scenarios involving the use of GenAI, which confirms its applicability under conditions of transformation of engineering practices [5]. The framework directs managers away from the search for a single universal metric in favor of a systemic analysis of engineering culture as a complex sociotechnical system.

A separate consideration is required for the epistemological status of DORA and SPACE metrics. DORA indicators primarily describe mechanical characteristics of the delivery pipeline and are by nature close to operational KPIs. In contrast, SPACE combines both strictly measurable and subjective parameters (for example, satisfaction and assessment of communication quality), forming a hybrid model that integrates quantitative and qualitative dimensions. This combination provides a foundation for constructing more complex statistical and sociotechnical models that make it possible to identify hidden dependencies between organizational structure, system architecture, and the actual outcomes of the release cycle [5, 9].

The implementation and operation of CI/CD pipelines remain incomplete without formalized service reliability management, as established in modern DevOps approaches and industry best practices [3]. Within the Site Reliability Engineering (SRE) framework, reliability is treated as a manageable parameter defined by target service quality indicators (SLOs) and measurable service level indicators (SLIs). Reliability thus ceases to be a vague characteristic and becomes the subject of formalized agreements between product owners, engineering teams, and service users.

The key mechanism for balancing innovation and stability is the concept of the error budget [6, 10]. This instrument establishes the permissible volume of deviations from SLOs in the form of a quantitatively measurable resource. As long as the actual values of incidents and degradations remain within the agreed-upon budget, it is acceptable to increase release intensity, experiment more actively with architecture, and accelerate the introduction of new features. When the budget is exhausted, a managerial shift occurs: the priority moves from introducing innovations to stabilizing the platform, eliminating defects, and preventing recurring incidents. As a result, a transparent and verifiable mechanism emerges for managing technological risks and prioritizing engineering efforts.

In practical terms, the error budget functions not only as a technical but also as a managerial regulator. At the level of organizational governance, a kind of social contract arises between product and SRE units: decisions about launching high-risk changes, postponing releases, or imposing moratoriums on deployments are made based on objective SLI/SLO metrics rather than subjective notions of acceptable risk. In highly regulated domains (financial sector, healthcare, critical infrastructure), this mechanism is integrated with legal and compliance requirements, enabling alignment of internal risk tolerance with external obligations concerning the availability, integrity, and security of services [3, 6, 10].

In scalable organizational systems, local optimization of CI/CD processes proves insufficient for ensuring long-term sustainability of outcomes [8]. The task arises of translating strategic guidelines into specific operational decisions made at the level of teams and value streams. Experience from Agile transformations shows that Lean Portfolio Management (LPM), as a component of the SAFe methodology, plays a significant role in addressing this task by ensuring alignment between strategy and execution [7].

LPM establishes a link between strategic priorities and the OKR system, on the one hand, and value streams on the other, using mechanisms of lean budgeting and portfolio governance [4]. Financing and restructuring of value streams cease to be episodic political-budgetary decisions and turn into a cyclical managerial process in which decisions rely, among other factors, on DORA metrics and reliability indicators. Thus, the strategy is translated into the language of release frequency, end-to-end delivery time, and service resilience, and portfolio hypotheses are validated through the actual dynamics of engineering metrics.

Under conditions of widespread adoption of GenAI solutions and the active use of cloud platforms, the portfolio level of management acquires an additional dimension. It becomes necessary to manage not only budget distribution across products but also the integral cost of change — the expenses associated with infrastructure, experimentation with models, controlled expansion of the ML landscape, and related risks [4, 7]. Integrating LPM with DevOps and SRE practices makes it possible to set threshold values for investments in experiments, link decisions about scaling specific streams to their demonstrated ability to deliver value consistently, and terminate funding of initiatives that fail to demonstrate the required effectiveness in a timely manner.

Empirical data obtained from scaling engineering organizations confirm the theoretical principles described above. In the author's professional experience, a case was implemented involving the growth of an engineering team from 40 to 120 specialists over approximately 1.5 years, management of a PMO function involving more than 50 managers, and a threefold expansion of the product portfolio — demonstrating the effectiveness of the Continuous Process Experimentation approach. The increase in organizational complexity was accompanied not by a one-time process reengineering effort but by a sequential series of controlled changes with explicit formulation of hypotheses and success criteria.

Short improvement cycles (A/B testing of process solutions, tuning and revising WIP limits in Kanban systems, modification of rules and depth of Code Review) were evaluated through the lens of DORA metrics and additional indicators reflecting the quality of collaboration and engineering team satisfaction. Empirically, a direct relationship was identified between successful scaling and the presence of a talent management contour — stable practices for managing competencies, career trajectories, and workload — combined with mature operational governance. Delivery of a large-scale product with DAU ~500k and revenue above 150M+ required a formalized architectural strategy in which technical debt and risk management, including ML components, were integrated into the main delivery pipeline and aligned with portfolio management decisions [6].

Additional analysis showed that the sustainability of scaling was determined not only by the maturity of CI/CD practices but also by the presence of institutionalized forums for developing architectural and product decisions — architectural committees, portfolio councils, regular retrospectives, and technical reviews. These mechanisms created the infrastructure for meaningful reversals — stopping or revising initiatives upon identifying systemic defects in process design, architecture, or management models. Ultimately, continuous process experimentation functioned not as a set of disconnected initiatives but as a controlled, measurable, and reproducible instrument for evolving the engineering organization [4, 6, 8].

From a methodological standpoint, the construction of longitudinal studies integrating DORA and SPACE metrics, SLO/SLI indicators, talent management data, and LPM portfolio-level decisions appears to be a promising direction. Such a research design will make it possible to identify causal relationships between process architecture, organizational structure, and the long-term scalability of technology businesses.

DISCUSSION

Based on the analysis of sources [1, 2, 9], a model is constructed that integrates existing process metrics (DORA) and team state metrics (SPACE). The conceptual core of this construct is that the dynamics of DORA indicators are subject to mandatory cross-validation through SPACE indicators. If, for example, an increase in deployment frequency or a reduction in time to restore service after incidents is observed while components of satisfaction within the SPACE model simultaneously decline, the system is classified as structurally unstable, being in a zone of elevated risk of burnout, latent accumulation of technical debt, and deterioration in the quality of managerial and engineering decision-making.

To represent this approach, an integrated management model scheme is used, in which the strategic goals of the organization are sequentially decomposed to the level of tactical and operational tasks, translated into the CI/CD pipeline, and the aggregate of metrics serves as a feedback channel for two planes of managerial influence at once: the operational one (management of development and operations streams) and the portfolio one (management of the set of products, initiatives, and resource allocation).

The proposed two-loop metrics architecture correlates with the basic principles of corporate governance, including the balance of stakeholder interests, the prevention of opportunistic behavior, and the reduction of agency costs. The DORA loop, aimed at measuring process performance and reliability, forms a quantitatively verifiable basis for managerial decisions, whereas the SPACE loop makes it possible to assess the resilience of human and organizational capital and to capture early indicators of crisis trends in team dynamics. Taken together, this creates conditions for a well-founded redistribution of authority, resources, and areas of responsibility within the organizational structure.

An additional dimension of the significance of the two-loop model is manifested in the sphere of normative and regulatory formalization of internal processes. Embedding DORA and SPACE metrics into the system of local regulatory acts, procedures, and risk management policies sets objective criteria for assessing the effectiveness of the IT function and digital products. This, in turn, reduces the conflictuality of strategic and budget planning, simplifies internal and external audit procedures, and increases the transparency of managerial decisions for corporate oversight bodies. Thus, two-loop efficiency measurement can be considered an instrument of not only technological but also legal and organizational support for the sustainable development of digital projects.

Figure 1 shows the integrated model: from strategy to stable releases [1, 2, 5].

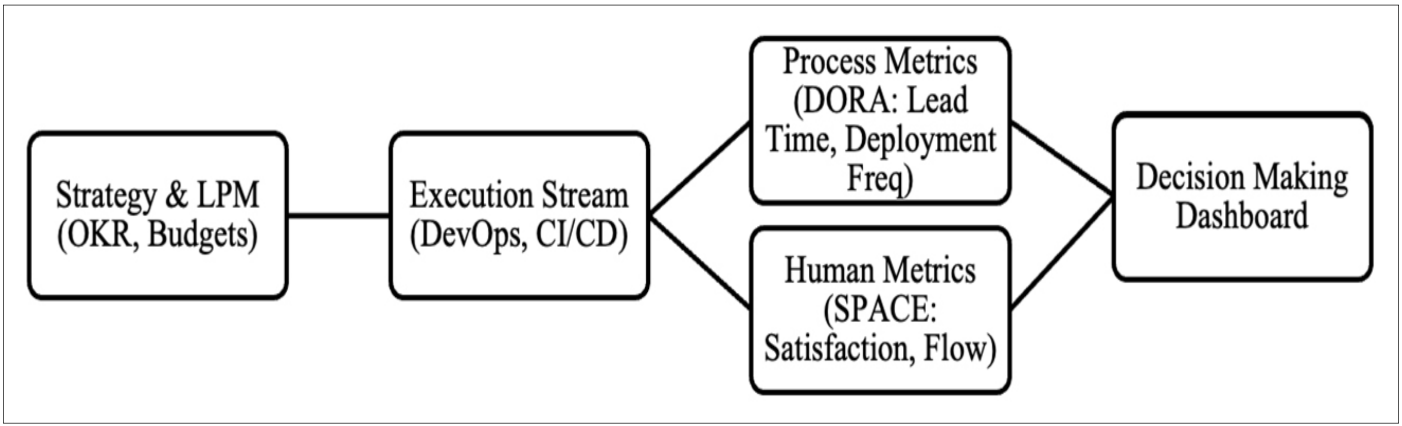

Fig.1. Integrated Model: From Strategy to Stable Releases [1, 2, 5]

The developed design model is aimed at reducing the likelihood of situations of partial, point optimization, in which technical personnel seek to demonstrate growth in quantitative indicators at the expense of weakening the functional and substantive components of the product being created. Reliance on integral rather than exclusively numerical performance criteria ensures the priority of the holistic quality of the resulting solution over the formal increase of individual metrics. A comparison of this approach with classical schemes, which rely predominantly on the step-by-step improvement of isolated indicators, is presented in Table 1.

Table 1. Comparative analysis of approaches to the evaluation of R&D effectiveness [4, 7]

Comparison criterion	Traditional KPIs (Waterfall)	Agile / Scrum metrics	DORA + SPACE + LPM (Author's approach)
Measurement focus	Resource utilization, hours	Velocity, Burn-down	Outcome, stability, well-being
Planning horizon	Long-term (years)	Sprint	Continuous flow (Flow)
Response to failure	Search for culprits, penalties	Retrospective	Error Budget Policy (system-wide stop)
Link to business	Weak, via specifications	Via Product Owner	Direct, via Value Streams and LPM

A special role in the architecture of the proposed model is assigned to the Error Budget mechanism. This instrument functions not so much as yet another numerical indicator as in the form of an embedded limiting loop for the CI/CD pipeline, which restrains excessive acceleration of the rate of change in the system and introduces disciplinary constraints into the release process. The need for a formalized definition and strict observance of the boundaries of the permissible level of failures and service degradation is elaborated in detail in SRE studies [6, 10], which emphasize that it is the Error Budget that ensures the institutionalized reconciliation of the interests of the evolutionary development of the product and the requirements for its reliability within a unified management system. The adoption of management decisions based on reliability indicators in this context can be described, in the terminology of system dynamics, as a balancing loop in which any impact on the development and operations loop is inevitably associated with the consumption or restoration of the error reserve.

From the standpoint of legal analysis of digital infrastructure, the Error Budget is capable of performing the function of an internal quasi-normative regulator that sets the de facto limits of acceptable technological risk at the organizational level. The incorporation of the corresponding threshold values into internal regulations, contracts for the provision of digital services and SLA makes it possible to transform abstract requirements for service reliability and continuity into legally significant criteria of proper performance of obligations. In this configuration, exceeding the established Error Budget limits confers an objectified character on the moment at which liability arises and serves as a basis for differentiating legal consequences: from adjustment of internal change management procedures to the application of contractual measures of influence in relations with customers and partners.

As engineering and product units scale, the Error Budget mechanism acquires fundamentally critical importance. Empirical data obtained in the course of expanding the development department from 40 to 120 specialists indicate that, in the absence of a formally established Error Budget operating regime, a multiple increase in team size is accompanied by a nonlinear, in fact exponential, growth in the number of incidents and a persistent degradation of key reliability indicators. In this regard, Table 2 contains a systematized set of metrics used at various stages of organizational growth and thereby reflects the evolution of requirements for managing the reliability and resilience of services as the structure and processes become more complex.

Table 2. Evolution of metrics depending on the scale of the organization [2, 7]

Scale (people)	Key focus	Main metrics	Management tools
Startup	Speed, Product-Market Fit	Deployment Frequency, Lead Time	Kanban board, CI/CD scripts
Scale-up	Stability, process hygiene	+ Change Failure Rate, MTTR	DORA dashboards, Error Budgets
Enterprise	Portfolio efficiency, alignment	+ Flow Efficiency, SPACE	LPM, SAFe, Portfolio Canvas

Finally, the implementation of the concept of continuous process experiments requires a clear decision-making algorithm for changes in CI/CD processes.

Thus, the formulated hypothesis is supported by empirical evidence: the management of modern software development objectively requires abandoning reliance on isolated one-dimensional metrics and transitioning to the use of a balanced system of indicators. Within such a system, the set of DORA indicators characterizes the production circuit, namely the pace and predictability of delivering changes, the SPACE model captures intra-team interaction and the state of professional well-being of specialists, whereas Lean Portfolio Management (LPM) sets the strategic development trajectory, ensuring the alignment of product initiatives with the organization's long-term guiding objectives.

The resulting triad of metrics institutionalizes a multilevel framework of managerial and, potentially, legal control over the life cycle of software products. Embedding DORA, SPACE, and LPM into the system of local regulatory acts, project activity regulations, as well as into contractual constructions (SLA, service level agreements, internal policies in the field of digital transformation) transforms abstract requirements for the efficiency and resilience of development into verifiable criteria of due diligence and proper performance of obligations. In this way, a methodological framework is formed for delineating the responsibility of different managerial levels for the outcome of digital projects, as well as for substantiated judicial and administrative control in disputes related to the quality and reliability of technological solutions.

CONCLUSION

As a result of the conducted study, a comprehensive model for assessing the effectiveness and resilience of software continuous integration and delivery processes has been developed and theoretically substantiated. It has been demonstrated that DORA metrics serve as a necessary but not sufficient criterion for the reliable diagnosis of large-scale engineering ecosystems: their isolated use inevitably leads to a systematic shift of emphasis toward the extreme acceleration of the pace of releases, accompanied by the simultaneous degradation of source code quality and deterioration of the professional and psychological condition of the involved specialists.

Embedding the SPACE framework into the observability and monitoring contour of CI/CD processes makes it possible to identify latent organizational and communication dysfunctions (including inter-team interaction barriers, chronic overload of holders of critical competencies, as well as a decrease in the subjective sense of controllability over development processes) that are fundamentally not represented on standard CI/CD dashboards, but have a direct and cumulative effect on the long-term productivity and resilience of engineering practice. The integration of SRE practices and, in particular, the Error Budgets mechanism with strategic portfolio management (LPM) tools creates an interpretive context for the analysis of metrics, transforming them from a predominantly punitive and controlling instrument into a means of managerial navigation and rational decision-making. The proposed dual-loop measurement model, in combination with the developed balancing algorithms (based on SLO and decision matrices), can be used by technical directors and other managers to design transparent, predictable, and scalable software delivery processes.

Of particular interest is the potential of the proposed model in the domain of regulatory and contractual formalization of digital processes. The formal consolidation of DORA metrics, SPACE indicators, Error Budget parameters, and LPM management practices in the internal local acts of an organization, corporate development standards, as well as in service level agreements (SLA), creates a basis for legally significant detailing of the parties' obligations and responsibilities for ensuring the reliability and predictability of change delivery. The incorporation of elements of dual-loop measurement into the system of internal control and audit makes it possible not only to represent the actual state of engineering practice, but also to serve as a criterion-based foundation for the allocation of responsibility in the event of incidents related to the quality and availability of digital services.

REFERENCES

1. Forsgren, N., Storey, M. A., Maddila, C., Zimmermann, T., Houck, B., & Butler, J. (2021). The SPACE of Developer Productivity: There's more to it than you think, 19(1).
2. John, B. (2025). Accelerating Software Delivery: Optimizing DevOps Pipelines for Speed and Quality. 1-5.
3. LAPOINTE-BOISVERT, A., TRUDEL, S., & DESHARNAIS, J. M. (2024). Measurement In DevOps Standard: Proposed Improvements, 1-16.
4. Voggenreiter, M., Angermeir, F., Moyón, F., Schöpp, U., & Bonvin, P. (2024, April). Automated security findings management: A case study in industrial devops. In Proceedings of the 46th International Conference on Software Engineering: Software Engineering in Practice (pp. 312-322). https://doi.org/10.1145/3639477.3639744.
5. Rüegger, J., Kropp, M., Graf, S., & Anslow, C. (2024, September). Fully Automated DORA Metrics Measurement for Continuous Improvement. In Proceedings of the 2024 International Conference on Software and Systems Processes (pp. 36-45). https://doi.org/10.1145/3666015.3666020.
6. Alozie, C. E., Akerele, J. I., Kamau, E., & Myllynen, T. (2025). Fault tolerance in cloud environments: Techniques and best practices from site reliability engineering. International Journal of Engineering Research and Development, 21(2), 191-204.
7. SAFe STUDIO Lean Portfolio Management. Retrieved from: https://framework.scaledagile.com/lean-portfolio-management/?utm_source (date accessed: 20.10.2025).
8. Ugwueze, V. U., & Chukwunweike, J. N. (2024). Continuous integration and deployment strategies for streamlined DevOps in software engineering and application delivery. Int J Comput Appl Technol Res, 14(1), 1-24.
9. Gebrewold, Y., & Wirell, J. (2023). Challenges with Measuring Software Delivery Performance Metrics: A case study at a software organisation, 20-44.
10. Khanna, K. (2024). Leveraging User Behavior-Based Service Level Indicators and Objectives to Prevent and Mitigate Software Incidents. International Journal of Advanced Research in Engineering and Technology (IJARET), 15(3), 353-364.
11. DORA Accelerate State of DevOps. Retrieved from: https://dora.dev/research/2024/dora-report/2024-dora-accelerate-state-of-devops-report.pdf (date accessed: 20.10.2025).