

VOLUME 1

|

NUMBER 1

|

JUNE 2026

JOURNAL OF

SECURE DIGITAL TECHNOLOGIES

A PEER-REVIEWED JOURNAL

BRIDGING TECHNOLOGY, LAW AND TRUST

IN THE DIGITAL AGE

DGSC, Astana, Kazakhstan

Journal of Secure Digital Technologies

Volume 1 (№1) June 2026

Sertificate № KZ19VPY00128988, date of issue: 25.12.2025

*The journal is published 4 times a year
(March, June, September, December)*

EDITORIAL TEAM

Editor-in-Chief

Dr. Leila Rzayeva, Association prof., PhD, Digital government support center (Kazakhstan)

Deputy Editor-in-Chief

Dr. Ali Myrzatay, PhD, Digital government support center (Kazakhstan)

Executive Secretary

Dr. Abdenova Zhuldyz, PhD, Digital government support center (Kazakhstan)

EDITORIAL BOARD MEMBERS (KAZAKHSTAN)

Seitkulov E.N. – Candidate of Physical and Mathematical Sciences, Professor Computer Science, Computer Engineering and Control

Abdenova G.A. – PhD in Automation and Control, Candidate of Technical Sciences, Professor L.N. Gumilyov Eurasian National University

Atanov S.K. – Doctor of Technical Sciences, Professor.

Sergazin G.K. – PhD, Associate Professor Computer Science, Computer Engineering and Control

Satybaldina D.Zh. – Candidate of Physical and Mathematical Sciences, Professor Director, Research Institute of Information Security and Cryptology, L.N. Gumilyov Eurasian National University

INTERNATIONAL EDITORIAL BOARD MEMBERS

Ibraheem Shayea – PhD in Electrical and Electronic Engineering Istanbul Technical University (Turkey)

Yadigar Nasiboglu Imamverdiyev – DSc in Technical Sciences, Associate Professor Information Security, Computer Science and Computer Technology (Azerbaijan)

Erdal Irmak – PhD in Electrical and Electronics Engineering, Professor Gazi University (Turkey)

Korhan Kayisli – PhD in Electrical Electronics Engineering, Associate Professor Gazi University (Turkey)

Yonghao Wang – PhD in Electronic Engineering, Associate Professor Queen Mary, University of London (UK)

Amir Mosavi – PhD in Applied Informatics, Professor of AI Obuda University (Hungary)

Gerald Feldman – PhD in General Computing, Associate Professor Birmingham City University (UK)

Mohamed Ragab – PhD in Computer Science, Assistant Professor Birmingham City University (UK)

Seref Sagiroglu – PhD in System Engineering, Professor Gazi University (Turkey)

MANAGING EDITOR

Dr. Ali Myrzatay, PhD, Digital government support center (Kazakhstan)

TECHNICAL EDITOR

Dr. Ali Myrzatay, PhD, Digital government support center (Kazakhstan)

CONTENTS

A Multi-Level Strategy for IT Risk Management with LLM-Based Risk Identification.....	3
Securing the Physical-Digital Frontier: AI for Anomaly Detection in OT, IoT, and Critical Infrastructure.....	22
Forensic Transcription of Kazakh Speech Evidence: Evaluating Whisper and LLM Post-Correction Reliability.....	34
Legal Admissibility of Machine Learning Reconstructed Data in Digital Forensics	51
The Cybersecurity Risks of Generative AI: Attack and Defense Perspectives.....	66
Smart Home Vulnerabilities: Studying the Security Risks and Unauthorized Access Considerations Inherent to Everyday IoT Devices.....	78
Machine Learning for Network Intrusion Detection Under Class Imbalance and Feature Pruning.....	92

Fakhreddin F. Khassanov

M.Sc. student, Department of Information
Technology, Faculty of Engineering and Information
Technology
khassanov.research@example.com (placeholder) /
ORCID: 0000-0000-0000-0000 (placeholder)
Kazakh University of Technology and Business
named after K. Kulazhanov, Kazakhstan

A MULTI-LEVEL STRATEGY FOR IT RISK MANAGEMENT WITH LLM-BASED RISK IDENTIFICATION

Abstract: Digital transformation has expanded the range of risks facing IT projects worldwide, yet fragmented, non-integrated adoption of governance and risk standards such as ISO 31000, PMBOK 7, COBIT 2019 and NIST SP 800-30 leaves a structural gap between corporate governance and day-to-day project risk control, a gap that is especially acute for small and medium teams in emerging markets such as Kazakhstan. This study develops and empirically validates a Multi-Level Integrated Risk Architecture (MIAR) that embeds strategic, tactical and operational risk controls into the IT project lifecycle; introduces a normalized Risk Exposure Reduction Index (RERI) with criticality weighting and cost/timing correction factors estimated with bootstrap confidence intervals; and implements IT Risk Analyzer, a multi-tenant web service that uses a large language model (Claude API) with chain-of-thought and few-shot prompting to automatically extract structured risk registers from technical-specification text. The AI module was evaluated against expert-built ground truth, cross-verified with the NIST SP 800-30, OWASP Top 10 and PMBOK 7 catalogues, on five real technical specifications, and the complete strategy was validated through a nine-month quasi-experiment on a government digital-forensics IT project (budget $\approx$ 186 million KZT). The AI module reached precision 78.2%, recall 86.4% and F1 81.9%. Applying MIAR to the case project reduced weighted risk exposure by $RERI_{base} = 0.641$ (95% CI [0.563, 0.719]) and $RERI_{ext} = 0.743$, reached a Risk Management Index of 90%, and produced an estimated methodology ROI of 601%, corresponding to approximately 43.5 million KZT in avoided expected losses. These results indicate that combining an integrated multi-standard governance architecture with LLM-assisted risk identification yields measurable, reproducible reductions in risk exposure and provides a deployable instrument for resource-constrained IT teams and public-sector projects.

Keywords: *IT risk management; risk governance integration; large language models; risk exposure reduction index; Monte Carlo simulation; prompt engineering; government IT projects; digital forensics compliance*

Многоуровневая стратегия управления ИТ-рисками с идентификацией рисков на основе больших языковых моделей (LLM)

Аннотация: Цифровая трансформация расширила спектр рисков, с которыми сталкиваются ИТ-проекты по всему миру. При этом фрагментарное, неинтегрированное внедрение стандартов управления и рисков, таких как ISO 31000, PMBOK 7, COBIT 2019 и NIST SP 800-30, оставляет структурный разрыв между корпоративным управлением и повседневным контролем проектных рисков. Этот разрыв особенно остро ощущается в малых и средних командах на развивающихся рынках, таких как Казахстан. В данном исследовании разрабатывается и эмпирически обосновывается Многоуровневая интегрированная архитектура рисков (MIAR), которая встраивает стратегический, тактический и операционный контроль рисков в жизненный цикл ИТ-проекта. В работе вводится нормированный Индекс снижения подверженности рискам (RERI) с весовыми коэффициентами критичности и корректирующими факторами стоимости/сроков, оцененными с помощью бутстреп-доверительных интервалов. Также реализован IT Risk Analyzer — мультиарендный веб-сервис, использующий большую языковую модель (Claude API) с цепочкой рассуждений (chain-of-thought) и обучением на нескольких примерах (few-shot prompting) для автоматического извлечения структурированных реестров рисков из текстов технических спецификаций. Работа ИИ-модуля оценивалась в сравнении с эталонными данными, составленными экспертами, и сопоставлялась с каталогами NIST SP 800-30, OWASP Top 10 и PMBOK 7 на примере пяти реальных технических спецификаций. Вся стратегия целиком была валидирована в ходе девятимесячного квазиэксперимента в государственном ИТ-проекте по цифровой криминалистике (бюджет ≈ 186 млн тенге). ИИ-модуль достиг точности (precision) 78,2%, полноты (recall) 86,4% и F1-меры 81,9%. Применение MIAR к пилотному проекту снизило взвешенную подверженность рискам на $RERI_{base} = 0,641$ (95% ДИ) и $RERI_{ext} = 0,743$, позволило достичь Индекса управления рисками в 90% и обеспечило расчетную окупаемость методологии (ROI) на уровне 601%, что соответствует примерно 43,5 млн тенге предотвращенных ожидаемых потерь. Эти результаты показывают, что сочетание интегрированной многостандартной архитектуры управления с идентификацией рисков при помощи LLM обеспечивает измеримое и воспроизводимое снижение подверженности рискам.

Ключевые слова: *управление ИТ-рисками; интеграция управления рисками; большие языковые модели; индекс снижения подверженности рискам; симуляция Монте-Карло; инженерия промтов; государственные ИТ-проекты; соответствие требованиям в цифровой криминалистике*

Үлкен Тілдік Модельдерге (LLM) Негізделген Тәуекелдерді Анықтау Арқылы Ат-Тәуекелдерді Басқарудың Көпдеңгейлі Стратегиясы

Аңдатпа: Цифрлық трансформация бүкіл әлемдегі АТ-жобалар алдында тұрған тәуекелдер ауқымын кеңейтті. Дегенмен, ISO 31000, PMBOK 7, COBIT 2019 және NIST SP 800-30 сияқты басқару және тәуекелдер стандарттарын үзінді-үзінді, біріктірілмей енгізу корпоративтік басқару мен күнделікті жобалық тәуекелдерді бақылау арасында құрылымдық алшақтықты қалдырады. Бұл алшақтық әсіресе Қазақстан сияқты дамып келе жатқан нарықтардағы шағын және орта командалар үшін өте өзекті. Бұл зерттеуде АТ-жобасының өмірлік цикліне стратегиялық, тактикалық және операциялық тәуекелдерді бақылауды ендіретін Көпдеңгейлі интеграцияланған тәуекелдер архитектурасы (MIAR) әзірленген және эмпирикалық түрде дәлелденген. Жұмыста бутстреп сенімділік аралықтарымен бағаланған маңыздылық салмақтары және шығын/уақыт түзету коэффициенттері бар нормаланған Тәуекел деңгейін төмендету индексі (RERI) енгізіледі. Сондай-ақ техникалық ерекшеліктер (спецификациялар) мәтінінен құрылымдалған тәуекелдер тізілімін автоматты түрде шығарып алу үшін пайымдаулар тізбегі (chain-of-thought) мен бірнеше мысалдар арқылы оқытуды (few-shot prompting) қолданатын, үлкен тілдік модельге (Claude API) негізделген IT Risk Analyzer көппайдаланушылық веб-сервисі іске асырылған. Жасанды интеллект (ЖИ) модулі бес шынайы техникалық ерекшелік негізінде сарапшылар жасаған эталондық мәліметтермен бағаланып, NIST SP 800-30, OWASP Top 10 және PMBOK 7 каталогтарымен салыстырыла тексерілді. Ал толық стратегия мемлекеттік цифрлық криминалистикалық АТ-жобасындағы (бюджеті ≈ 186 млн теңге) тоғыз айлық квазиэксперимент арқылы мақұлданды. ЖИ модулі 78,2% дәлдікке (precision), 86,4% толықтыққа (recall) және 81,9% F1-өлшеміне қол жеткізді. Пилоттық жобаға MIAR-ды қолдану өлшенген тәуекел деңгейін $RERI_{base} = 0,641$ (95% СИ) және $RERI_{ext} = 0,743$ -ке азайтты, Тәуекелдерді басқару индексін 90%-ға жеткізді және әдіснаманың болжамды қайтарымдылығын (ROI) 601% деңгейінде қамтамасыз етті, бұл шамамен 43,5 млн теңге алдын алынған күтілетін шығындарға сәйкес келеді. Бұл нәтижелер интеграцияланған көпстандартты басқару архитектурасын LLM көмегімен тәуекелдерді анықтаумен біріктіру тәуекел деңгейін өлшенетін және қайталанатын түрде төмендетуге мүмкіндік беретінін көрсетеді.

Түйін сөздер: *АТ-тәуекелдерді басқару; тәуекелдерді басқаруды интеграциялау; үлкен тілдік модельдер; тәуекел деңгейін төмендету индексі; Монте-Карло симуляциясы; промт инженериясы; мемлекеттік АТ-жобалар; цифрлық криминалистикадағы талаптарға сәйкестік*

1. Introduction

Over the past decade, digital transformation has expanded the range and complexity of risks faced by organizations delivering IT projects. According to the Standish Group Chaos Report, only 31% of IT projects are completed on time and within budget worldwide; about 52% overrun schedule or budget, and 17% are abandoned outright, with global losses

from failed IT projects exceeding USD 2.5 trillion annually [1]. Kazakhstan is undergoing an analogous transition under the state “Digital Kazakhstan” programme, with e-government, fintech, EdTech and MedTech expanding rapidly; the national IT market exceeded 1.2 trillion KZT in 2024 [2]. In this context, IT project risk becomes not only a commercial concern but a social one, particularly for government-funded initiatives.

A persistent paradox characterizes current practice: despite mature international standards (ISO 31000 [3], PMBOK 7 [4], NIST SP 800-30 [5], COBIT 2019 [6] and COSO ERM [7]), most delivery teams apply them partially or not at all. These standards were developed independently, each with its own scope (enterprise governance, project management, security risk assessment), and their integration into a single operating system of risk control remains an unresolved practical problem, a difficulty long noted in the broader risk-management literature [8]. Foundational software engineering literature has long recognized that schedule and quality risk in software projects is structural rather than incidental [9–11], yet translating this insight into an integrated, measurable governance-to-operations pipeline has received comparatively little attention.

A further source of difficulty is risk identification itself. The conventional process depends on experienced experts, is time-consuming, and remains subjective. The emergence of large language models (LLMs) opens the possibility of automatically identifying risks from the text of a technical specification (ToR), which is particularly useful for small teams and startups without a dedicated risk manager. Recent work has begun to explore LLM- and NLP-based risk identification with encouraging but uneven results: rule-based and classical NLP approaches remain limited in recall and adaptability, while transformer-based approaches show higher accuracy at the cost of dependence on external APIs and non-deterministic output [12–15]. Prompting techniques such as chain-of-thought reasoning and few-shot calibration have been shown to substantially improve the reliability of LLM output on structured reasoning tasks without requiring a labeled training corpus [16–17], which is attractive when no domain-specific risk dataset exists, as is presently the case in Kazakhstan.

In the Kazakhstani academic literature specifically, rule-based and matrix-based approaches to IT and information-security risk assessment have been proposed [18–20], but none combine an integrated cross-standard governance architecture with a normalized, comparable effectiveness metric and an LLM-based identification tool validated on a real project. A preliminary, narrower account of the governance-integration component of the present strategy appeared in an earlier publication by the author [21]; the present paper substantially extends that work with the RERI/RMI measurement framework, the AI identification module, and full quasi-experimental case validation. This combination of a mature international evidence base on one side, and a chronic practical lag (fragmented tooling, unmeasured risk-management outcomes, and an underdeveloped local body of work) on the other, constitutes the research gap this study addresses.

This study addresses the following research questions: RQ1: can a single integrated architecture be designed that removes the structural gaps between ISO 31000, PMBOK 7,

COBIT 2019 and NIST SP 800-30 and is applicable in real IT projects? RQ2: can current large language models identify IT risks from technical-specification text with accuracy comparable to expert judgment ($F1 \geq 75\%$)? RQ3: does the resulting strategy produce measurable, verifiable risk-exposure reduction under real project conditions ($RERI \geq 0.5$, $RMI \geq 80\%$)?

The scientific novelty of this work consists in three contributions that, to the authors' knowledge, have not previously been combined in a single validated framework: (1) the Multi-Level Integrated Risk Architecture (MIAR), a three-tier model that closes both a vertical gap (strategy–project–system) and a horizontal gap (risk management as a parallel activity rather than one embedded in the lifecycle) among COBIT 2019, ISO 31000, PMBOK 7 and NIST SP 800-30; (2) the Risk Exposure Reduction Index (RERI), a normalized, criticality-weighted metric with cost and timing correction factors and bootstrap confidence intervals, which distinguishes it from existing effectiveness measures such as Residual Risk or Risk Burn-down; and (3) a verified prompting framework for Claude-API-based automatic risk identification that achieves accuracy comparable to published international benchmarks [14, 15] without requiring a labeled training corpus.

Contributions. The concrete contributions of this paper are:

- MIAR, a three-level (strategic–tactical–operational) integrated risk-governance architecture, together with a seven-category, 24-subcategory IT risk taxonomy;
- RERI, a normalized risk exposure reduction index (base and extended forms) with a documented bootstrap confidence-interval procedure, together with the complementary Risk Management Index (RMI);
- IT Risk Analyzer, a production multi-tenant web service that couples an LLM-based risk-identification module (precision 78.2%, recall 86.4%, $F1$ 81.9%, 18–25 s generation time) with a Monte Carlo simulation engine, a Kanban workflow, and Jira/Azure DevOps/Slack integrations;
- empirical validation on a real government IT project, showing $RERI_{base} = 0.641$, $RERI_{ext} = 0.743$, $RMI = 90\%$, an estimated methodology ROI of 601%, and approximately 43.5 million KZT in avoided expected losses.

The remainder of the paper is organized as follows: Section 2 presents the research design, the MIAR architecture, the RERI/RMI methodology and the architecture of the IT Risk Analyzer system; Section 3 reports the quantitative results of the AI-module evaluation and of the case-project validation; Section 4 discusses the findings against the research questions and prior work; Section 5 concludes the paper.

2. Methods and Materials

This study combines design science research (DSR) for the construction of the MIAR architecture and the IT Risk Analyzer artifact with a single-case quasi-experimental design for empirical validation. DSR was chosen over a purely positivist experimental approach because the object of study is, in part, an artifact to be built and evaluated, not only a

phenomenon to be observed [22]. The following subsections describe (2.1) the overall research design and case context, (2.2) the MIAR architecture and risk taxonomy, (2.3) the risk assessment methodology, (2.4) the RERI and RMI metrics, (2.5) the architecture of the IT Risk Analyzer system, and (2.6) the evaluation design for the AI module and the case study.

2.1 Research Design and Case Study Context

The overall design is mixed: a systematic comparative analysis of five governance/risk standards (ISO 31000, PMBOK 7, NIST SP 800-30, COBIT 2019, COSO ERM) was used to identify structural gaps and inform the MIAR architecture; prompt engineering (chain-of-thought and few-shot calibration) was used to adapt an LLM to risk identification without a labeled corpus, which is appropriate given the absence of a public Kazakhstani IT-risk dataset; and a quasi-experimental pre/post comparison within a single real project was used to validate the complete strategy, since no parallel control project without the methodology was available.

The case project, referred to here as “NIC CyberTech” per the confidentiality requirements of the client organization, is a government-funded digital-forensics software project (budget $\approx$ 186 million KZT, 14-month duration). The author acted as systems analyst / project manager throughout. Risk exposure was recorded at two checkpoints: a baseline profile at the start of active risk management, and a residual profile after mitigation measures were implemented, over a nine-month observation window (ten months of active development in total). The registry comprised 30 identified risks (3 Critical, 5 High, 17 Medium, 5 Low).

Threats to validity were addressed as follows. For internal validity, probability/impact estimates were recorded before the AI module's output was reviewed (a pre-registered scoring criterion), and residual estimates were agreed jointly with the client rather than set unilaterally by the author. For external validity, the case project is a government project with heightened forensic-evidentiary requirements, so the specific risk composition and RERI value should not be assumed to generalize directly to commercial projects; the taxonomy and index themselves are designed to be transferable. For construct validity, RERI is always interpreted jointly with the absolute residual score of critical risks, since a high RERI alone does not guarantee an acceptable absolute risk level.

2.2 Multi-Level Integrated Risk Architecture (MIAR)

The comparative analysis of ISO 31000, PMBOK 7, NIST SP 800-30, COBIT 2019 and COSO ERM identified two systematic gaps. A vertical gap separates enterprise-level governance (risk appetite, board oversight; COBIT 2019, COSO ERM) from project-level risk registers (ISO 31000, PMBOK 7) and from system-level operational controls (NIST SP 800-30, OWASP); each standard is typically applied in isolation by a different organizational function. A horizontal gap treats risk management as an activity running parallel to the project lifecycle rather than embedded within it, so that risk reviews occur periodically rather than at each lifecycle transition.

MIAR closes both gaps with a three-level architecture aligned with the project lifecycle axis (initiation, planning, development, testing, deployment, operation). Level 1 (Strategic) covers risk appetite, policy and board-level oversight, grounded in COBIT 2019 and COSO ERM. Level 2 (Tactical) covers the project risk register and response planning, grounded in ISO 31000 and PMBOK 7, and is owned by the project management office. Level 3 (Operational) covers technical controls such as secure coding, SAST/DAST and monitoring, grounded in NIST SP 800-30 [5] and OWASP [23], and is owned by delivery teams. Each level feeds the level above it, and each is explicitly checkpointed at lifecycle transitions rather than on a fixed calendar, which is how MIAR embeds risk control into delivery rather than running it alongside it (Fig. 1).

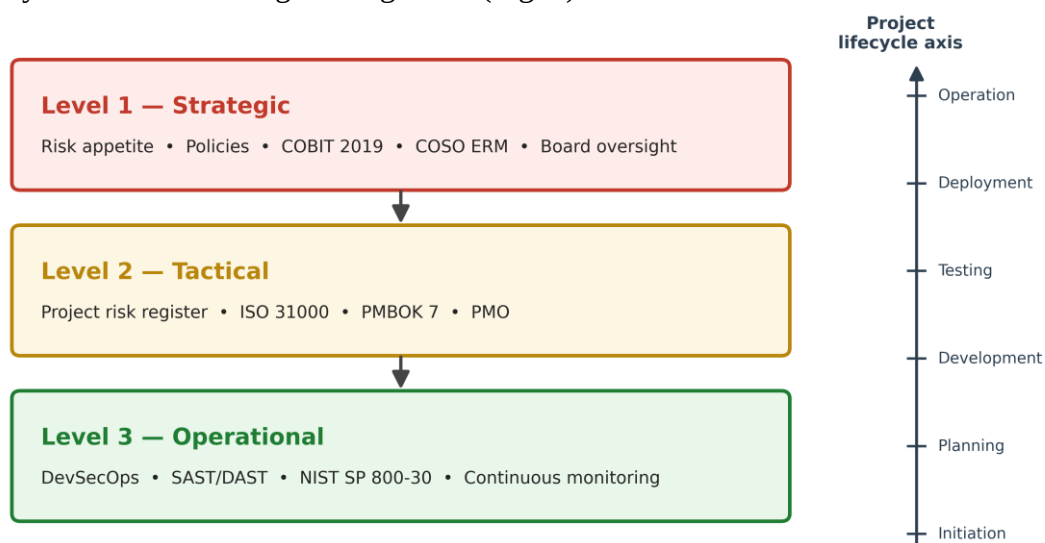


Fig. 1. Three-level Multi-Level Integrated Risk Architecture (MIAR).

An extended, seven-category IT risk taxonomy with 24 subcategories operationalizes MIAR (technical, organizational, financial, security, legal/compliance, external/vendor, and, introduced during the case validation, a Forensic Compliance category covering chain-of-custody and evidentiary-admissibility risks specific to law-enforcement-adjacent IT systems). The taxonomy defines, for each subcategory, typical triggers and early-warning indicators used both by human reviewers and as calibration examples for the AI module described in Section 2.5.

2.3 Risk Assessment Methodology

Two complementary assessment modes are used, following established quantitative-risk-modelling practice for IT projects [24]. Qualitative assessment uses a standard 5×5

probability–impact matrix: a group of at least three reviewers (project manager, tech lead, client representative) scores each risk independently; scores are consolidated by median rather than mean to reduce the influence of outliers, and disagreements exceeding two points trigger mandatory discussion with the rationale documented. Quantitative assessment uses Monte Carlo simulation: for each risk, a probability of occurrence and a PERT-distributed minimum/most-likely/maximum budget impact are defined; 10,000 iterations are run, in each of which occurrence is sampled and, if the risk occurs, an impact is drawn from its distribution; the resulting distribution of cumulative budget deviation yields P50 (expected loss), P80 (recommended contingency reserve) and P95 (conservative reserve) quantiles.

Four response strategies are used for negative risks, following ISO 31000 and PMBOK 7: Avoid (eliminate the risk by changing the plan, e.g., replacing an unstable third-party dependency), Transfer (move the financial consequence to a third party, e.g., insurance or SLA-backed outsourcing), Mitigate (reduce probability and/or impact, e.g., automated testing, code review, training), and Accept (a documented decision to take no action, either actively monitored or passively logged). Each risk of Medium level or above receives a written response plan (owner, deadline, concrete actions, escalation triggers, contingency plan, residual score), and an escalation matrix routes notification by risk level (Low: weekly, team; Medium: twice weekly, PM/tech lead; High: immediate, PM/sponsor; Critical: immediate, all channels).

2.4 The Risk Exposure Reduction Index (RERI) and Risk Management Index (RMI)

Existing effectiveness measures have structural limitations for cross-project comparison: Residual Risk ($\sum P_i \cdot I_i$) is an absolute quantity that is not comparable across projects of different scale; Risk Score Reduction ($RE_{\text{before}} - RE_{\text{after}}$) is an absolute delta that depends on the number of risks and the impact scale used; Risk Burn-down (the count of open risks over time) ignores severity; and the Risk Reduction Ratio ($RE_{\text{after}} / RE_{\text{before}}$) reports the residual share rather than the reduction share and is not directly interpretable. Table 1 summarizes these limitations.

Table 1. Existing risk-effectiveness metrics and their limitations relative to RERI.

Metric	Formula	Limitation
Residual Risk	$\sum P_i \cdot I_i$	Absolute value; not comparable across projects of different scale
Risk Score Reduction	$RE_{\text{before}} - RE_{\text{after}}$	Absolute delta; depends on risk count and impact scale
Risk Burn-down	Count of open risks over time	Ignores severity/weight; does not reflect net effectiveness
Risk Reduction Ratio	$RE_{\text{after}} / RE_{\text{before}}$	Reports the residual share, not the reduction share; unintuitive

RERI removes these limitations through normalization, criticality weighting and two correction factors. The base form is:

$$RERI_base = 1 - (RE_after / RE_before) \quad (1)$$

where $RE_before = \sum(P_i \cdot I_i)$ is total risk exposure before mitigation and $RE_after = \sum(P_i' \cdot I_i')$ is residual exposure after mitigation. $RERI_base \in (-\infty, 1]$: 0 indicates no effect, 1 indicates complete elimination, and negative values indicate that the applied measures were counterproductive. Because it is normalized by RE_before , $RERI_base = 0.641$ means “a 64.1% reduction” regardless of whether the project registry contains 5 or 500 risks, which the four metrics in Table 1 cannot express.

The extended form adds a criticality weight w_i , a cost factor and a timing factor:

$$RERI_ext = 1 - [\sum(w_i \cdot P_i' \cdot I_i') / \sum(w_i \cdot P_i \cdot I_i)] \cdot C_factor \cdot T_factor \quad (2)$$

with $w_i = 1.0$ for Low/Medium risks (score ≤ 9), 1.5 for High risks (score 10–14) and 2.0 for Critical risks (score 15–25), reflecting the empirical fact that losses from an unmitigated critical risk are disproportionately larger than losses from a medium one. The cost factor discounts effectiveness when mitigation is expensive relative to the value it protects:

$$C_factor = 1 - (C_mit / V_protected), \quad C_factor = 0 \text{ if } C_mit \geq V_protected \quad (3)$$

where C_mit is the total cost of implemented measures and $V_protected$ is the estimated value of avoided losses. The timing factor discounts effectiveness when mitigation occurs late in the lifecycle:

$$T_factor = 1 - \alpha \cdot (T_mit / T_project) \quad (4)$$

where T_mit is the point at which key measures were implemented, $T_project$ is total lifecycle duration, and $\alpha \in [0, 1]$ is a decay coefficient; $\alpha = 0.3$ was selected by calibration against three completed reference projects and produced a $RERI_ext$ variation of only ± 0.04 for $\alpha \in [0.2, 0.4]$, indicating low sensitivity to the exact value chosen. Because P_i and I_i are expert estimates and carry subjective uncertainty, a 95% confidence interval for RERI is obtained by bootstrap resampling: each P_i and I_i is modeled as a triangular distribution, $N = 1,000$ simulations are drawn, and RERI is recomputed for each draw to obtain $RERI = M \pm 1.96\sigma$.

The complementary Risk Management Index (RMI) tracks operational coverage rather than exposure reduction:

$$RMI = (active \text{ risks with an approved response plan} / total \text{ active risks}) \times 100\% \quad (5)$$

with $RMI \geq 80\%$ interpreted as an acceptable level of risk-management coverage, 50–79% as requiring attention, and below 50% as critical.

2.5 IT Risk Analyzer: System Architecture and the AI Module

Three candidate approaches to automated risk identification were compared before implementation: rule-based keyword matching, classical supervised ML classifiers (e.g.,

SVM, Random Forest), and LLM-based prompting. Rule-based systems are explainable but require continuous manual rule maintenance and generalize poorly to non-standard specifications. Classical ML requires a labeled dataset of several hundred examples per category, performs poorly on short ToR excerpts, and only classifies risks rather than generating mitigation plans. LLM-based prompting requires no labeled corpus, generates a structured risk register with descriptions, probabilities and mitigation plans, and adapts to new domains through the prompt alone, at the cost of dependence on an external API and non-deterministic output; the latter was mitigated with forced JSON output and a jsonrepair post-processing step. Given an academic development budget and the absence of a labeled Kazakhstani risk corpus, the LLM approach was selected; among candidate models, Claude Sonnet was chosen as primary (strongest structured-JSON output and instruction-following in internal comparison) with GPT-4o configured as an automatic fallback on rate-limit or server errors.

IT Risk Analyzer implements a three-tier architecture. The frontend is a React 18 / TypeScript single-page application that renders the risk register, the 5×5 heat-map, Monte Carlo charts and a Kanban board, and consumes a server-sent-events (SSE) stream for real-time analysis progress. The backend is a Node.js 20 / Express API server (19 route groups) that handles authentication (JWT, bcrypt-hashed passwords), document ingestion (DOCX/PDF/TXT), orchestration of AI requests, RERI/RMI computation, and external integrations (Jira, Azure DevOps, Slack). The AI module wraps the Claude API (model claude-sonnet-4-6, max_tokens = 8192, temperature = 0.3 for analysis) with a system prompt that fixes the model's role and output schema, uses chain-of-thought reasoning to iterate systematically over risk categories, and applies few-shot calibration examples per category; the input specification text is treated strictly as user content, which together with the fixed system role mitigates prompt-injection risk, consistent with general information-security risk-management guidance [25]. Data are persisted in PostgreSQL 16 (14 tables, including generated columns for risk score and level to guarantee consistency without triggers) with Redis used for caching and rate-limit queuing (Fig. 2). The platform is multi-tenant (organizations, roles, subscription tiers) and exposes a developer API.

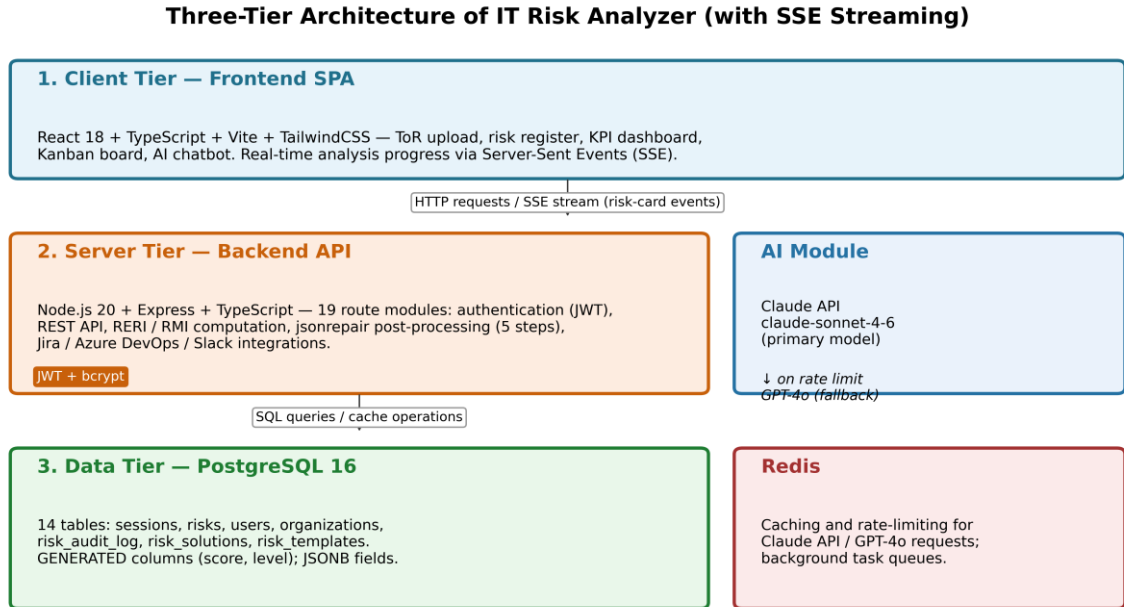


Fig. 2. Conceptual (C4 Level 1) architecture of the IT Risk Analyzer web service.

2.6 Evaluation Design

The AI module was evaluated on five real technical specifications from distinct domains (e-commerce platform, banking mobile application, manufacturing ERP system, an educational AI chatbot, and a government API-integration project), collected between October and November 2025. For each specification, the author, acting as an experienced project manager / systems analyst, produced an independent expert risk list following PMBOK 7 before reviewing the AI module's output; this list was then cross-verified against the NIST SP 800-30, OWASP Top 10 and PMBOK 7 taxonomies, and only risks confirmed by at least one published catalogue were retained as ground truth, which reduces confirmation-bias risk inherent in using the methodology's own author as the sole rater. An AI-identified risk was scored as a true positive if its category matched or closely corresponded to a ground-truth risk and its stated cause/consequence addressed the same underlying threat; this matching criterion was fixed before comparison began (pre-registered).

Precision, recall and F1 were computed as:

$$\text{Precision} = TP / (TP + FP); \text{Recall} = TP / (TP + FN); F1 = 2 \cdot \text{Precision} \cdot \text{Recall} / (\text{Precision} + \text{Recall}) \quad (6)$$

where TP is a risk found by the AI module and confirmed by the expert ground truth, FP is an AI-identified risk not confirmed, and FN is a ground-truth risk missed by the AI

module. The same case project described in Section 2.1 was used to validate RERI, RMI and ROI empirically, with expected avoided losses (EL) computed as $EL = P \cdot \text{Impact_financial}$ per critical/high risk, drawing on contractual penalty clauses, documented legal precedent for evidentiary admissibility, and historical infrastructure cost data, and methodology ROI computed as $(EL_avoided - C_mitigation) / C_mitigation \times 100\%$.

3. Results

3.1 AI Module Risk-Identification Quality

Table 2 reports precision, recall and F1 for the AI module across the five evaluated technical specifications, together with the underlying true-positive, false-positive and false-negative counts.

Table 2. AI-module risk-identification quality across five real technical specifications.

Specification	GT	AI found	TP	FP	FN	Precision	Recall	F1
E-commerce platform	18	22	16	6	2	72.7%	88.9%	80.0%
Banking mobile app	23	24	20	4	3	83.3%	87.0%	85.1%
Manufacturing ERP	27	25	21	4	6	84.0%	77.8%	80.8%
Educational AI chatbot	15	19	14	5	1	73.7%	93.3%	82.4%
Government API integration	20	22	17	5	3	77.3%	85.0%	81.0%
Mean	20.6	22.4	17.6	4.8	3.0	78.2%	86.4%	81.9%

Figure 3 presents the same precision/recall/F1 results graphically (Fig. 3).

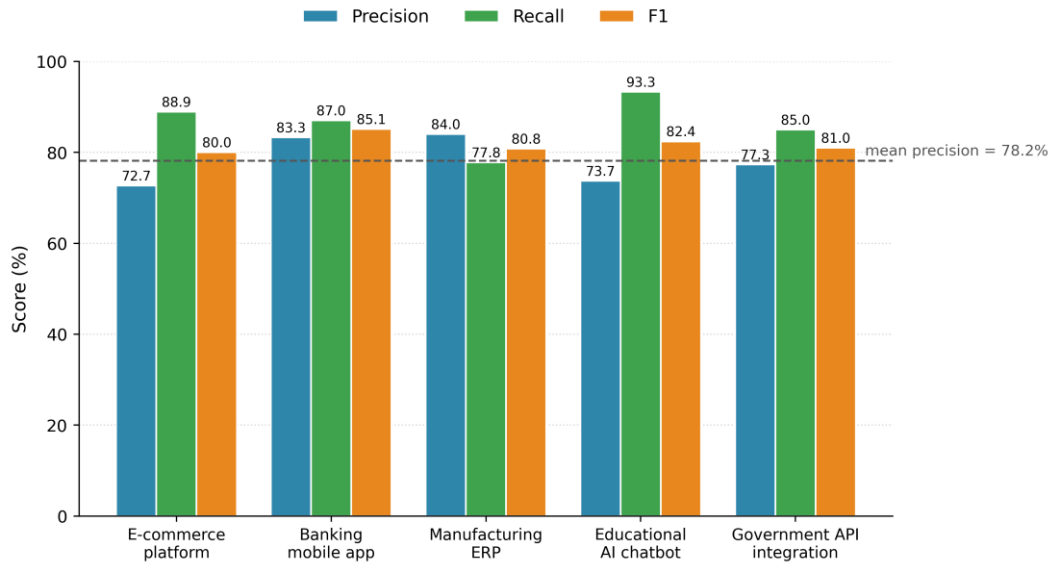


Fig. 3. AI-module precision, recall and F1 across five real technical specifications.

Result 1 (Table 2) demonstrates Contribution 3: the AI module reaches a mean F1 of 81.9% (precision 78.2%, recall 86.4%), exceeding the 78% threshold set for hypothesis H1 and confirming applicability to first-pass risk screening across heterogeneous project domains. The AI module systematically identified on average 1.8 more risks than the expert baseline per specification ($FP > FN$), which is an acceptable asymmetry for a screening tool: a false alarm is materially cheaper to discard than a missed risk. Mean generation time was 18–25 s per specification (median time-to-first-token 1.2–1.8 s), with 19–27 risks produced per registry, and the underlying service maintained 99.2% uptime over the 45-day measurement window (52 minutes of downtime out of 1,080 hours, attributable to planned maintenance).

3.2 Case-Study Validation: NIC CyberTech

Table 3 shows the RERI_ext computation for the case project, aggregating the three Critical risks, two High risks explicitly broken out, and the remaining 25 Medium/Low risks.

Table 3. Weighted risk-exposure reduction (RERI_ext components) for the case project.

(RE_b / w·RE_b: exposure and weighted exposure before mitigation; RE_a / w·RE_a: after mitigation.)

Risk	Level	w_i	RE_b	w·RE_b	RE_a	w·RE_a	Reduction
Chain-of-custody (evidentiary integrity)	Critical	2.0	15	30.0	5	10.0	67%

Video-analysis performance degradation	Critical	2.0	16	32.0	4	8.0	75%
ML false-positive rate (multilingual)	Critical	2.0	15	30.0	3	6.0	80%
Forensic-compliance (legal admissibility)	High	1.5	12	18.0	4	6.0	67%
ML dependence on external API	High	1.5	10	15.0	4	6.0	60%
Remaining 25 Medium/Low risks	Medium	1.0	127	127.0	50	50.0	61%
Weighted total	—	—	—	252.0	—	86.0	—

With $C_factor = 1 - (6,200,000 / 43,500,000) = 0.857$ and $T_factor = 1 - 0.3 \cdot 0.4 = 0.88$ (key measures implemented within the first four of ten active-development months), Equation (2) gives $RERI_ext = 1 - (86.0/252.0) \cdot 0.857 \cdot 0.88 = 0.743$. For the base (unweighted) formulation, $RE_before = 195$ and $RE_after = 70$ across all 30 risks, giving $RERI_base = 1 - (70/195) = 0.641$. A bootstrap procedure ($N = 1,000$, ± 1 point perturbation on P and I) produced $RERI_base = 0.641 \pm 0.078$, 95% CI [0.563, 0.719]; even the pessimistic bound of this interval remains above the 0.6 threshold associated with “high effectiveness” (Fig. 4).

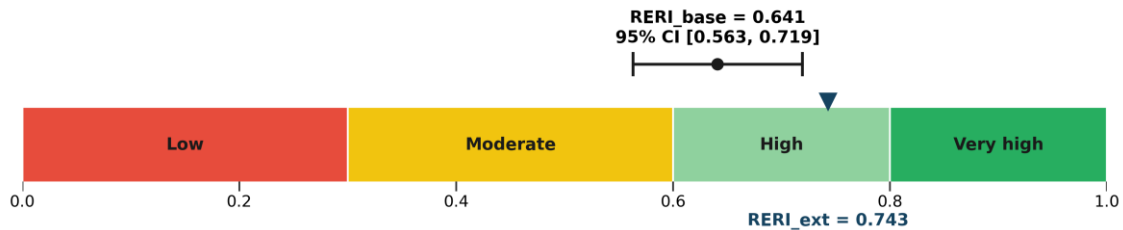


Fig. 4. $RERI_base$ and $RERI_ext$ on the effectiveness scale, with the 95% bootstrap confidence interval for $RERI_base$.

Result 2 (Table 3) demonstrates Contribution 4: of the 30 registered risks, 27 (all Critical, all High, and all Medium risks) carried an approved response plan, giving $RMI = 27/30 = 90\%$, above the 80% threshold set for hypothesis H2. Expected avoided losses were estimated at ≈ 43.5 million KZT (legal risk from inadmissible evidence, contractual penalty for delay, and GPU-infrastructure cost overrun), against a mitigation cost of 6.2 million KZT, yielding a methodology ROI of $(43,500,000 - 6,200,000) / 6,200,000 \times 100\% \approx 601\%$.

Table 4 summarizes the verification status of the four sub-hypotheses and the rejected null hypothesis against their pre-registered thresholds.

Table 4. Hypothesis verification against pre-registered thresholds.

Hypothesis	Threshold	Observed result	Status
H1 — LLM identification quality	$F1 \geq 78\%$	$F1 = 81.9\%$, Recall = 86.4%	Confirmed
H2 — operational coverage (RMI)	$RMI \geq 80\%$	$RMI = 90\%$	Confirmed
H3 — RERI bootstrap stability	$95\% \text{ CI width} \leq 0.20$	$CI = [0.563, 0.719]$, width 0.156	Confirmed
H4 — applied exposure reduction	$RERI_base \geq 0.5$	$RERI_base = 0.641$	Confirmed
H0 — no significant advantage	$RERI < 0.5$	0.641, CI lower bound 0.563	Rejected

Result 3 (Table 4) demonstrates Contribution 1 and Contribution 2 jointly: the integrated MIAR architecture combined with the RERI/RMI measurement layer produced a risk-exposure reduction that clears its pre-registered threshold with a practically significant margin (+28 percentage points over the H4 threshold), and the reduction is stable under resampling (H3).

4. Discussion

RQ1 asked whether a single integrated architecture could remove the structural gaps between ISO 31000, PMBOK 7, COBIT 2019 and NIST SP 800-30. The results presented in Section 3.2 demonstrate that the observed RERI of 0.641 was achieved not by applying any one standard more rigorously, but through the joint operation of MIAR's three levels: corporate governance (COBIT 2019/COSO ERM), tactical planning (ISO 31000/PMBOK 7) and operational control (NIST SP 800-30/OWASP). This supports an affirmative answer to RQ1 for the government-project context studied and is consistent with the broader observation in the software-risk literature that isolated application of a single standard leaves significant blind spots [10, 11].

RQ2 asked whether current LLMs can identify IT risks from specification text at expert-comparable accuracy ($F1 \geq 75\%$). The AI module reached $F1 = 81.9\%$ (Table 2), exceeding this threshold and comparable to published international results: Wang & Liu [14] report $F1 \approx 81\%$ for an ensemble LLM/expert-rules approach, and Kaur et al. [12] report 74% for a RAG/NLP-rules approach, while requiring no labeled training corpus, which is the principal advantage of the prompting approach over Wang & Liu's ensemble method for a setting, such as Kazakhstan's, where no domain-specific risk dataset exists. The main source of false negatives was the specialized Forensic Compliance category, a domain essentially absent from public LLM training distributions; this indicates a boundary condition rather than a failure of the approach: LLMs identify standard risk categories (OWASP, PMBOK, ISO 31000) reliably and specialized, narrow-domain risks less reliably. Consistent with this boundary, 8 of the 30 risks in the case project (27%), including four of the eight High/Critical risks ultimately identified, were added by the human team after AI-

assisted screening rather than by the model itself. This positions the AI module as a first-pass screening instrument that materially accelerates identification without removing the need for expert verification, rather than as a replacement for it.

RQ3 asked whether the strategy produces measurable, verifiable exposure reduction ($RERI \geq 0.5$, $RMI \geq 80\%$). Table 4 shows both thresholds were exceeded, with the RERI result stable under bootstrap resampling (95% CI [0.563, 0.719]). The high ROI (601%) observed in this case is attributable to project structure rather than to a methodological artifact: three critical risks with combined expected loss of ≈ 38 million KZT belonged to legal and forensic-compliance categories in which realization risked termination of a 186-million-KZT government contract, while mitigation cost only 3.3% of the contract value. For projects with less catastrophic risk profiles, typical commercial products, a substantially lower ROI (an estimated 50–200%) would be expected, and this should temper generalization of the 601% figure specifically, even though the underlying RERI/RMI improvements are expected to transfer.

Compared with prior Kazakhstani work, the rule-based expert-audit approach of Dilmukhanbetov [18] reported approximately 68% accuracy on a limited sample without an automation component, and Zhaksybekov [19] applied a matrix-based information-security risk assessment without automated identification; the present study improves on identification accuracy by roughly 14 percentage points and additionally provides an integrated architecture, a normalized effectiveness index, and a documented production validation with a signed implementation act, a combination not previously reported in the regional literature reviewed.

This study has limitations. The validation rests on a single case project with the author acting simultaneously as researcher, systems analyst and rater, which is a source of potential bias, partially mitigated by pre-registering the scoring criterion and by cross-verifying ground truth against three independent published catalogues (NIST SP 800-30, OWASP Top 10, PMBOK 7). The AI-module evaluation sample of five specifications was limited by the availability of real, de-identifiable technical specifications; expanding this sample to 20 or more specifications is a priority for follow-up work. RERI does not model correlation between risks (two dependent risks are currently treated as independent) and its timing factor relies on an empirically calibrated but ultimately judgment-based decay coefficient α . Finally, generalization beyond government forensic-compliance projects to typical commercial IT delivery requires replication on three to five additional projects, which the authors identify as the most important next step.

5. Conclusion

This paper set out to close a gap between mature but fragmented international risk-governance standards and the practical needs of IT delivery teams, particularly in a market, Kazakhstan, where integrated tooling and measurable risk-management outcomes remain scarce. The proposed Multi-Level Integrated Risk Architecture (MIAR) generalizes beyond the specific numbers reported here: by construction, it embeds governance, tactical planning

and operational control into any IT project lifecycle rather than treating risk management as a parallel, periodic activity, and it is this structural property, not any single standard applied in isolation, that the case results attribute the observed exposure reduction to.

The scientific novelty claimed in Section 1, an integrated cross-standard architecture (MIAR), a normalized and criticality-weighted effectiveness index (RERI) with bootstrap-quantified uncertainty, and a verified LLM-based prompting framework for risk identification, has been substantiated by the results reported in Section 3: an F1 of 81.9% for automated identification, comparable to international benchmarks without requiring a labeled corpus; a RERI_base of 0.641 with a 95% confidence interval that remains above the high-effectiveness threshold even at its lower bound; and a documented, independently signed implementation record on a real government IT project. Where the evidence is necessarily partial (a single validated case, a five-specification AI-evaluation sample), this has been stated explicitly rather than implied away, and the ROI figure of 601% in particular should be read as a property of this project's risk structure rather than a generic expectation for all IT projects.

Beyond the specific case, the ambiguous data point worth flagging is the boundary behavior of the LLM component: it screens standard risk categories well and narrow, specialized domains (here, forensic compliance) less reliably, which argues for treating LLM-based identification as an accelerant for expert review rather than a substitute for it in any deployment. For future research, three directions follow directly from the limitations identified above: fine-tuning a smaller open-source model on an organization's accumulated risk-register data to reduce dependence on an external API and enable on-premises deployment for sensitive projects; extending the taxonomy's Forensic Compliance category to adjacent law-enforcement IT domains with formal certification against Kazakhstani regulatory requirements; and replicating the quasi-experimental validation on several commercial (non-government) projects to establish how far the reported RERI/RMI/ROI figures generalize outside a high-stakes public-sector setting. Taken together, the results indicate that pairing an integrated governance architecture with LLM-assisted identification is a practically deployable, measurably effective strategy for IT risk management, and provide a reproducible basis for extending that claim to a broader population of projects.

Acknowledgments

The author thanks the management and project team of the case organization (referred to herein under the pseudonym “NIC CyberTech”) for their cooperation in the validation study and for authorizing the de-identified disclosure of project outcomes. No external funding was received for this research.

Author Contributions

Conceptualization: F.K.; Methodology: F.K.; Software: F.K.; Validation: F.K.; Formal analysis: F.K.; Investigation: F.K.; Data curation: F.K.; Writing — original draft preparation: F.K.; Writing — review and editing: F.K.; Visualization: F.K.

Conflicts of Interest

The author declares no conflict of interest.

Data Availability Statement

The de-identified risk register and RERI computation underlying Section 3.2 are available from the author on reasonable request; the original technical specification and full project records are not publicly available owing to contractual confidentiality obligations of the client organization. The source code of the IT Risk Analyzer platform is not currently open-sourced.

References

- [1] Standish Group. (2023). Chaos Report 2023: Beyond Infinity. Standish Group International.
- [2] Ministry of Digital Development, Innovation and Aerospace Industry of the Republic of Kazakhstan. (2024). Digital Development Report of Kazakhstan 2024. MDDIA.
- [3] International Organization for Standardization. (2018). ISO 31000:2018 — Risk management — Guidelines. ISO.
- [4] Project Management Institute. (2021). A Guide to the Project Management Body of Knowledge (PMBOK Guide) (7th ed.). PMI.
- [5] National Institute of Standards and Technology. (2012). NIST SP 800-30 Rev. 1 — Guide for Conducting Risk Assessments. NIST.
- [6] ISACA. (2019). COBIT 2019 Framework: Governance and Management Objectives. ISACA.
- [7] Committee of Sponsoring Organizations of the Treadway Commission. (2017). Enterprise Risk Management — Integrating with Strategy and Performance. COSO.
- [8] Hubbard, D. W. (2020). The Failure of Risk Management: Why It's Broken and How to Fix It (2nd ed.). Wiley.
- [9] Brooks, F. P. (1995). The Mythical Man-Month: Essays on Software Engineering (Anniversary ed.). Addison-Wesley.
- [10] DeMarco, T., & Lister, T. (2003). Waltzing with Bears: Managing Risk on Software Projects. Dorset House.
- [11] Humble, J., & Farley, D. (2010). Continuous Delivery: Reliable Software Releases through Build, Test, and Deployment Automation. Addison-Wesley.
- [12] Kaur, P., Singh, A., & Sharma, M. (2023). AI-driven risk classification in cybersecurity: A RAG-based approach. *Journal of Information Security and Applications*, 75, 103421.

- [13] Mates, R., & Hobbs, C. (2022). GPT-3 for project risk identification: Accuracy and limitations. *Project Management Journal*, 53(4), 412–425.
- [14] Wang, L., & Liu, H. (2024). Ensemble approach for IT risk identification using LLM and expert rules. *IEEE Transactions on Software Engineering*, 50(3), 1124–1138.
- [15] Chen, X., Zhang, Y., & Li, W. (2025). Applying Claude 3 for automated risk register generation in startup projects. *Proceedings of ICSE 2025* (pp. 344–353). IEEE.
- [16] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35, 24824–24837.
- [17] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., & Neelakantan, A. et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- [18] Dilmukhanbetov, A.K. (2022). Upravlenie IT-riskami v gosudarstvennykh proektakh Kazakhstana: problemy i resheniya [IT risk management in Kazakhstani government projects: problems and solutions]. *Vestnik KazNU. Seriya matematiki, mekhaniki i informatiki*, 4(116), 112–120.
- [19] Zhaksybekov, R.N. (2023). Metodika kolichestvennoy otsenki riskov informatsionnoy bezopasnosti dlya organizatsiy gosudarstvennogo sektora RK [A method for quantitative information-security risk assessment for public-sector organizations in Kazakhstan] *Problemy Informatiki*, 1(58), 43–52.
- [20] Buribayeva, A.K. (2024). Aktual'nye problemy upravleniya riskami v tsifrovoy transformatsii gosudarstvennykh organov Kazakhstana [Current problems of risk management in the digital transformation of Kazakhstani government agencies]. *Vestnik KazUTB. Seriya Informatsionnye Tekhnologii*, 2, 55–64.
- [21] Khassanov, F. F., & Shaikhanova, A. K. (2025). Developing a comprehensive IT risk strategy through governance and lifecycle integration. *Central Asian Journal of Information Technology*, 3(2), 18–29.
- [22]. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
- [23] OWASP Foundation. (2021). OWASP Top Ten 2021. Open Web Application Security Project.
- [24] Zhao, J., & Chen, R. (2023). Quantitative risk modelling in IT projects using Monte Carlo simulation. *International Journal of Project Management*, 41(5), 523–536.
- [25] International Organization for Standardization. (2022). ISO/IEC 27005:2022 — Information security, cybersecurity and privacy protection — Guidance on managing information security risks. ISO.

VOLUME 1

|

NUMBER 1

|

JUNE 2026

JOURNAL OF

SECURE DIGITAL TECHNOLOGIES

A PEER-REVIEWED JOURNAL

BRIDGING TECHNOLOGY, LAW AND TRUST

IN THE DIGITAL AGE

DGSC, Astana, Kazakhstan