

A Four Stage Evidence Based Assessment Protocol for Student Work with Large Language Model Assistance and Offline Verification of Rubric Scoring and Critical Error Gates

Nayyar Ahmed Khan^{1*}, Mohammed Ahmad² & Rim Hamdaoui²

¹Department of Computer Engineering, College of Computing and Information Technology, Shaqra University, Saudi Arabia.

²Department of Computer Science, College of Science and Humanities, Dawadmi, Shaqra University, Saudi Arabia.

Corresponding Author (Nayyar Ahmed Khan) Email: nayyar@su.edu.sa*



DOI: <https://doi.org/10.46759/iijsr.2026.10309>

Copyright © 2026 Nayyar Ahmed Khan et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Article Received: 18 July 2026

Article Accepted: 23 September 2026

Article Published: 30 September 2026

ABSTRACT

This paper defines a four-stage assessment protocol for bounded large language model (LLM) assistance, with initial work, optional feedback, checked revision, and explanation. A weighted rubric and independent critical-error gate were tested on six constructed evidence vectors. Identical vectors receiving artificial intelligence (AI) and peer feedback both scored 90.0, while a trace with a critical source error scored 62.5 and was held for review. The arithmetic test verifies scoring behavior, not marker reliability or learning. The paper specifies anchors, calibration, accessibility routes, and a future transfer evaluation.

Keywords: Generative Artificial Intelligence; Authentic Assessment; Process Evidence; Large Language Model Assistance; Rubric Scoring; Critical Error Gate; Source Verification; Marker Calibration; Assessment Fairness; Independent Transfer; Academic Integrity; Student Agency.

1. Introduction

A polished take-home product is an incomplete measure of learning when an LLM can brainstorm, edit, or generate substantial portions of it. A blanket ban may also exclude legitimate feedback or accessibility support. The technical assessment problem is to define the target skill, permit bounded assistance, and collect evidence that the learner can interpret sources and explain a consequential choice. This paper proposes a four-stage protocol applicable to a single assignment [1-6]. The design has two research questions for a future pilot: does the protocol provide more reliable evidence of the intended skill than a final product alone, and can learners complete it equitably without excessive workload? It does not assume that process evidence proves authorship. Instead, several complementary observations inform a teacher's judgment. Integrity decisions remain subject to institutional procedure [7-10].

The paper contributes an assistance matrix, an evidence record, rubric with anchored criteria, a sample scoring trace, and an evaluation plan. The example is synthetic and is included to demonstrate the method. No detector is used as the sole basis for an inference about a student.

1.1. Study Objectives

- 1) Specify a bounded assistance rule tied to the learning outcome.
- 2) Collect initial work, feedback, revision, and explanation evidence.
- 3) Define weighted rubric anchors and a separate critical-error gate.
- 4) Verify arithmetic and route invariance on constructed traces.
- 5) Provide equivalent feedback and explanation routes for access needs.
- 6) Plan a study of marker reliability, workload, and transfer.

2. Literature Review

Education discussion since 2020 has examined how generative AI changes writing, coding, and feedback tasks. A final answer may reflect student reasoning, tool output, feedback from others, or a mixture. Process-oriented assessment responds by making planning, revision, and explanation visible. It must remain proportional: exhaustive prompt logs can be intrusive and may measure compliance more than subject understanding [11-14]. Recent guidance on AI competency and human oversight also emphasizes critical verification and equitable access. Students differ in language background, disability-related support, connectivity, and willingness to use AI. An assessment should permit a comparable non-AI feedback route and grade the target skill consistently across routes. The present paper defines a protocol rather than claiming a measured advantage over conventional assignments.

For a future comparison, the outcome should include a later independent task requiring the same reasoning with different source content. The analysis must predefine whether it examines group mean scores, rubric reliability, or teacher workload as the primary endpoint. Prior achievement, task difficulty, and feedback access are plausible confounders. A crossover design could reduce some differences but may have carryover effects because students learn the process. The pilot should first establish feasibility, timing, and the clarity of the AI rule [15-18]. The scoring record should preserve criterion-level ratings, evidence locations, and a short marker rationale. For a source-based task, the evidence location can be a passage identifier and a line in the student's response. The record should not include a speculative "probability of AI authorship." A teacher can see why a rating was assigned and another marker can revisit it. A critical-error flag, such as a fabricated document quote, is stored separately from the weighted total. The policy defines its consequence before grading begins [19].

The process note has three fields: assistance type, consequential suggestion, and learner decision. A sample entry is "I requested counterarguments; I rejected one because the assigned source does not state that date; I revised my second claim using paragraph four." This is enough to direct the teacher to a checkable decision. It is not proof that the learner wrote every sentence [20-22]. The same format can describe feedback from a peer, keeping the evidence requirement comparable across AI and non-AI routes.

An evaluation compares the staged task with a similar task that uses only a final product, while accounting for instructional differences. Planned measures include marker agreement, time spent per student, student understanding of the assistance rule, access to both routes, and performance on a later independent transfer question [23-26]. Sampling, consent, missing-data handling, and an analysis plan should be fixed before outcome data are collected. The protocol is proposed; these measures have not been obtained.

Privacy and fairness rules limit the evidence collected. A process note records meaningful decisions, not every private prompt [27-30]. Teachers should not demand a paid account. Oral explanation has an equivalent format for a learner who cannot use it. A mismatch between stages invites clarification and additional evidence; it is not automatically misconduct. The school's formal integrity process governs consequential findings.

[1] found that authentic-looking outputs alone did not reliably protect assessment integrity in their experimental contexts. [2] argues for visible justification and responsibility within work supported by generative artificial

intelligence. The proposed protocol therefore records an initial claim, a consequential revision, and an explanation, then applies a critical-error gate independently of the weighted score. It does not treat the process record as proof of authorship.

3. Research methodology

Let O be the learning outcome, A the allowed assistance set, and E the evidence record. For a source-based writing task, O is the ability to compare claims using assigned documents; A may include brainstorming and clarity feedback but excludes invented evidence or outsourced final interpretation. E contains an initial claim, selected source passages, a brief revision note, the final response, and an explanation of one decision. The teacher publishes O and A before students begin.

Stage 1 is a short initial attempt under stated conditions. Stage 2 permits optional AI, peer, or teacher feedback. Stage 3 requires a revised submission and a compact note identifying a consequential suggestion, whether it was accepted, and why. Stage 4 asks the learner to explain one decision through a written, recorded, or live format. The same checkpoint types apply to students who choose a non-AI route. Full prompt histories are not required unless an established process specifically calls for them.

An assistance matrix describes what each tool may do. For example, an LLM may propose counterarguments, but the learner must check them against the assigned source. It may suggest clearer wording, but the learner must verify factual changes. In coding, it may suggest a test case, while the learner must run the program and explain the fix. The rule should include allowed and disallowed examples, a question channel, and a disclosure instruction that can be completed in a few sentences. Figure 1 summarizes the sequence and the authority boundary described above.



Figure 1. Four stage assessment evidence path. Initial reasoning is followed by optional feedback, a checked revision, and an explanation of a consequential decision; the same evidence stages apply to non-AI feedback. Source: Original schematic prepared for this manuscript.

The rubric separates four dimensions: evidence accuracy, reasoning, revision judgment, and communication. An illustrative weighting is 40, 30, 20, and 10 percent, respectively. For dimension j , a marker assigns an anchored score $s(j)$ from 0 to 4. The total is the sum of $\text{weight}(j)$ times $s(j)$ divided by 4. This calculation is transparent and can be adjusted before release to match another subject. A polished style cannot compensate for a fabricated source under an explicit critical-error rule.

Equation (1) states the acceptance or scoring rule used by the proposed method; its variables are defined in the surrounding text and Table 1.

$$\text{Score} = 100 \times (0.40 s_E + 0.30 s_R + 0.20 s_V + 0.10 s_C) / 4;$$

$$\text{CriticalError} \Rightarrow \text{review, independent of Score} \quad (1)$$

In Equation (1), Score is the weighted mark out of 100; s_E , s_R , s_V , and s_C are criterion ratings from 0 to 4 for evidence, reasoning, revision, and communication, respectively. The coefficients 0.40, 0.30, 0.20, and 0.10 sum to one, and division by 4 converts the maximum criterion rating to a fraction. CriticalError denotes an unresolved consequential factual error and routes the work to review regardless of Score.

Algorithm 1. Offline decision procedure

INPUT E0 initial claim, E1 feedback, E2 revision, E3 explanation

Mark evidence, reasoning, revision, communication from 0 to 4

Compute weighted score with published anchors

IF critical factual fabrication: hold for teacher review

Compare $E0 \rightarrow E2$ change with source and E3 explanation

Record criterion rationale; offer equivalent non-AI route

The evidence record can be represented as a directed sequence E0 to E3. E0 is the initial idea, E1 the feedback event, E2 the revised product, and E3 the learner's explanation. A revision edge records what changed and why. The teacher need not inspect every intermediate draft; one consequential change is sufficient for the core protocol. The design tests whether the explanation refers to the assigned evidence rather than to an unsupported model suggestion. A missing stage does not automatically imply misconduct; it is a prompt for clarification or an access check.

Rubric anchors should be written before student work is viewed. For evidence accuracy, the top anchor requires that claims match the assigned documents and that the link between passage and interpretation is explicit. A middle anchor may use a relevant passage but leave the inference partly unexplained [31-36]. A low anchor includes a material unsupported claim. Similar anchors define reasoning, revision judgment, and communication. Markers should also identify critical factual fabrication separately, because averaging can mask an error that changes the substantive conclusion.

An accessibility route should preserve the construct being assessed. A learner unable to give a live oral explanation can provide a recorded or written explanation of the same decision. A learner without reliable AI access can obtain comparable peer or teacher feedback. The marker scores the evidence judgment and reasoning with the same rubric. The school should document any difference in time or feedback availability between routes. This is an assessment-equivalence question to test in the pilot, not a claim that every route is identical [37].

For reliability analysis, each criterion should be examined separately. Exact agreement is the proportion of double-marked items assigned the same anchor; adjacent agreement counts ratings within one level. A chance-adjusted statistic may be appropriate for ordinal ratings, but it should be reported with the sample size and uncertainty. If markers disagree substantially, revise the anchor or train markers before treating the total score as stable [38]. The goal is defensible interpretation of learning, not a high statistic produced by collapsing categories. Table 1 defines the main variables and control checks used in this method.

Table 1. Assistance and evidence matrix distinguishing permitted brainstorming, source location, and clarity support from the learner's responsibility for interpretation.

Activity	Permitted assistance	Learner evidence
Brainstorm	Possible angles or questions	Initial and chosen claim
Source check	Candidate passage location	Original passage interpretation
Revision	Clarity feedback	Accepted or rejected change
Final explanation	No generated justification	Reason for a key decision

Calibration samples should be selected to challenge superficial scoring. Include a polished but unsupported interpretation, a grammatically rough yet well-evidenced comparison, and a revision that improves style while weakening accuracy. Markers score them independently, discuss disagreements, and revise anchors before the actual cohort is marked [39-42]. The training record identifies the final rubric version. If the rubric changes after grading starts, earlier work must be rescored or the change postponed to avoid inconsistent standards.

The transfer task can ask students to compare a new pair of short sources without AI assistance under stated conditions. It should assess the same reasoning operation but not reuse the exact facts or wording of the take-home assignment. The pilot can compare transfer performance across assessment designs while monitoring baseline skill and differential access. A prespecified missing-data rule is needed for absent learners [43]. Transfer is a stronger outcome than the attractiveness of the final submission, but it still needs careful interpretation.

A small arithmetic unit test contains six constructed evidence traces with rubric vectors (s_E, s_R, s_V, s_C), tool route, and a critical-error flag. The weighted score is calculated from the published weights. Two traces intentionally share the same vector while using AI and peer feedback, testing route invariance of the formula. One trace includes a critical source error, testing that the review gate operates independently of the numerical score. These traces are not human-marked student work and cannot estimate rubric reliability or learning [44-49].

Before grading, two markers independently score a small anonymized calibration set that includes fluent but weak reasoning and less polished but well-supported reasoning. They compare differences by criterion and revise ambiguous anchors. During a pilot, a subset of work is double marked. Report exact agreement, criterion-level disagreement, and the reason for adjudication. A single overall score can conceal whether the evidence criterion is unreliable.

4. Results and Discussion

The scoring function produced 90.0 for both the AI-feedback trace A and peer-feedback trace B because both had the vector (4,3,4,3). This is a property of the stated formula: the route is not an input to the score. Trace C, with vector (1,3,4,4), produced 62.5 and was held for review because it was marked with a critical source error. The other constructed totals were 77.5, 50.0, and 90.0. The calculated outputs are tabulated below. Table 2 records the constructed trace or issue case used to illustrate the decision path.

The critical-error gate is deliberately separate from the weighted score. An unsupported central fact cannot be neutralized by fluent communication or a polished process note. The teacher must inspect the original source and

the learner's explanation before deciding on feedback or any formal action. The synthetic test verifies the arithmetic and gate behavior, not whether the anchors are valid or markers apply them consistently. Table 3 reports the offline component-test outputs and their denominators.

Table 2. Constructed evidence record that links an initial claim, feedback, source-checked revision, and explanation to the teacher's assessment decision.

Stage	Recorded item	Teacher interpretation
Initial	Tentative similarity claim	Starting reasoning
Feedback	Three suggested counterarguments	Not graded as student work
Revision	Reject two unsupported claims	Source judgment visible
Explanation	Defend one passage choice	Probe independent understanding

Table 3. Weighted rubric outputs for six constructed traces; the critical-error gate is applied independently of the numerical score and no learner outcomes are inferred.

Trace	Feedback route	Scores E,R,V,C	Total	Gate
A	AI	4,3,4,3	90.0	Eligible
B	Peer	4,3,4,3	90.0	Eligible
C	AI	1,3,4,4	62.5	Review
D	None	3,4,2,3	77.5	Eligible
E	AI	2,2,1,4	50.0	Eligible
F	Peer	4,4,3,2	90.0	Eligible

The worked writing trace E0–E3 shows how a model suggestion can be rejected against a source and how the final explanation identifies a defensible claim. Its educational value remains a hypothesis. A real pilot must compare marker agreement, time, access, and independent transfer performance; none of those outcomes can be inferred from six constructed vectors.

The protocol may improve interpretability of a submission by showing a decision path, but it adds checkpoints and marking time. A future pilot must measure that cost. Students can also produce formulaic process notes, so the explanation should connect to a specific claim, source, or code change. The independent transfer task tests whether reasoning extends beyond the submitted product. It remains possible for a student to obtain outside help at every stage; no single design proves authorship.

Fairness depends on access to feedback and to an appropriate explanation format. A student using translation or accessibility support should not be penalized when those tools are allowed [50]. A detector score or prose “style” should not substitute for evidence of learning. If a concern arises, teachers should seek clarification through established procedures and interpret multiple pieces of evidence. Clear instructions before the deadline are part of fair assessment.

The approach can be adapted by discipline: a mathematics student explains an error correction, a scientist interprets a graph, and a programmer demonstrates a boundary test. The specific checkpoints and rubric anchors must follow

the target skill. The present source-based example cannot establish that the protocol works equally well in all subjects.

Process evidence can itself become performative. A learner might write a plausible note after the fact, and a teacher might overvalue a fluent explanation. The independent transfer task and calibration samples reduce, but do not eliminate, this problem. Markers should interpret E0 to E3 together with the actual sources and avoid treating one discrepancy as proof of improper tool use. The protocol is designed to improve the evidence available for judgment, not to create a surveillance record of every action [51].

Implementation cost is a material limitation. Staged checkpoints use class time; double marking uses staff time; alternate feedback routes require coordination. A pilot should measure these demands alongside any change in rubric reliability or student understanding [52-56]. If a checkpoint adds little information, remove it. The technical specification is deliberately modular so a discipline can retain only the stages that support its learning outcome while keeping transparent assistance rules.

Rubric reliability and validity are different. Two markers may agree closely on a criterion that does not represent the learning outcome. The assessment team should therefore inspect whether the checkpoints and anchors actually elicit source comparison. Student interviews and later transfer tasks can help assess that question. A high agreement statistic is useful only alongside meaningful task design and a fair opportunity to demonstrate the skill.

If the workflow becomes burdensome, the teacher can use fewer checkpoints while keeping the initial attempt, a meaningful revision, and a final explanation. The choice should be documented before students begin. A high-stakes context may require supervised individual performance as well [57]. The proposed protocol is most defensible when it is transparent, proportionate, and adapted to the discipline, rather than applied as a universal form to every assignment.

A validity threat arises when the process note becomes a second writing assignment. Students skilled at describing a revision may receive credit that should reflect source comparison, while students with strong reasoning may give brief explanations. The rubric should therefore score the decision supported by the actual work, not elegance of the note. Marker calibration can include examples of terse but adequate rationales. This protects the intended construct and reduces the risk that the assessment rewards compliance with a preferred narrative style.

The assessment conditions must be documented precisely. An initial in-class checkpoint can establish a starting idea, but it may be affected by time pressure, language, or access needs [58]. The teacher should explain how that checkpoint contributes to the score and offer an equivalent format where required. The final explanation should test a decision the student actually made; a generic question that can be answered by memorizing a rubric provides weaker evidence. These choices are part of the intervention and must be reported in any future study.

Comparing the staged task to a product-only task presents a measurement problem. The staged task generates more observable evidence, which may change marker confidence without changing student knowledge. Report confidence and agreement separately from learner performance. If a later independent task shows no difference, the protocol might still improve fairness or diagnostic feedback, but it should not be described as a learning

intervention on that basis alone [59]. The research questions and primary outcome need to be fixed before seeing results.

Institutional integrity procedures remain separate from routine marking. A mismatch between an explanation and a final product warrants a conversation, but neither a weak process note nor an automated detector establishes intent. Document the opportunity to clarify, preserve the relevant evidence, and apply the same procedure across students. A transparent assessment design can reduce ambiguity but cannot eliminate every dispute about assistance.

The synthetic trace can be evaluated at each edge. E0 shows a broad similarity claim. At E1, the model supplies one potentially useful distinction and two unsupported statements. E2 shows the learner rejecting unsupported statements and tying the retained distinction to a passage. E3 tests whether the learner can explain that source choice. The trace supports an inference about demonstrated source judgment in this example; it does not establish that the student independently generated every word of the final response.

An alternative failure trace is also useful. If E2 repeats the invented date and E3 cannot identify a passage for it, the evidence-accuracy criterion is weak and the teacher can request a correction or further explanation. The appropriate response follows the assessment and integrity rules; an automated detector is unnecessary for this initial pedagogical judgment. These traces illustrate how the same protocol handles a successful verification and a material source error [60-62].

The synthetic example produces a visible revision edge. The broad initial claim is narrowed after one defensible counterargument is checked against the assigned documents. The evidence record identifies the source passage and the rejected unsupported date. A marker could inspect whether the final interpretation actually follows from the passage and whether the learner's explanation reflects that choice. The model suggestion itself receives no credit; the learner's verification and reasoning do.

A contrasting example shows why a process note alone is insufficient. Suppose a learner writes that they rejected an invented date but the final response still includes it. The marker must assess the submitted claim and request clarification, rather than awarding revision credit for the note's wording. The staged protocol makes that inconsistency visible. It does not tell the teacher whether the error was deliberate, accidental, or introduced during later editing.

5. Conclusion and Future Recommendation

This paper specifies a four-stage assessment protocol that makes source judgment, revision, and explanation observable while allowing bounded AI assistance. The proposed rubric and evaluation plan support a testable classroom pilot. Claims about reliability, workload, or learning require actual data before the design is adopted broadly.

5.1. Future Recommendations

- 1) Test the rubric with independent human markers and report agreement.
- 2) Compare staged and product-only tasks on later transfer questions.

- 3) Examine fairness across language and accessibility support routes.
- 4) Measure the time burden of each evidence checkpoint.
- 5) Adapt the protocol to coding, mathematics, and laboratory work.
- 6) Study whether process notes reveal reasoning or become formulaic.

Declarations

Source of Funding

The authors received no funding for this contribution.

Competing Interests Statement

There are no conflicts of interest by the authors in any way.

Consent for publication

The authors give consent for the publication.

Authors' contributions

All the authors contributed equally to the article.

Informed Consent

Not Applicable. The reported offline checks used constructed records and no human participants.

Availability of data and material

Constructed test inputs and deterministic procedures are described in the manuscript.

Institutional Review Board Statement

Not Applicable. No human participants or identifiable student records were included in the offline checks.

Ethical Approval

Not Applicable to the constructed offline unit tests.

References

- [1] Kofinas, A.K., Tsay, C.-H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British Journal of Educational Technology*, 56(6): 2522–2549. <https://doi.org/10.1111/bjet.13585>.
- [2] Sharma, S. (2026). Educational integrity in GenAI-augmented assessment: Making judgement visible. *International Journal for Educational Integrity*, 22: 6. <https://doi.org/10.1007/s40979-026-00216-y>.
- [3] Ahmad, M., et al. (2025). Learning three-dimensional face recognition from sparse views for robust identity verification. Available at SSRN 5428214.

- [4] Akram, F., et al. (2024). Integrating artificial bee colony algorithms for deep learning model optimization: A comprehensive review. In *Solving with Bees: Transformative Applications of Artificial Bee Colony Algorithm*, Pages 73–102.
- [5] Al Omari, O.M.A., Khan, N.A., & Mahafdah, R. (2017). Ranking and reputation based resource allocation in P2P system. *Mediterranean Journal of Basic and Applied Sciences*, 1(1): 293–301.
- [6] Alanezi, R., Alanezi, M.A., & Khan, N.A. (2018). Development of web based E-cooperative training system. In *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, IEEE.
- [7] Alangari, S., et al. (2022). Developing a blockchain-based digitally secured model for the educational sector in Saudi Arabia toward digital transformation. *PeerJ Computer Science*, 8: e1120.
- [8] Alangari, S., & Khan, N.A. (2021). Artificially intelligent warehouse management system. *Asian Journal of Basic Science & Research*, 3(3): 16–24.
- [9] Alotaibi, R., Khan, N.A., & Akhtar, M.M. (2025). Statistical probability prediction model for E-learning and realtime proctoring using IoT devices. *Journal of King Saud University–Science*, 37.
- [10] Alshalaan, M., & Khan, N.A. (2025). Complexities and challenges for securing digital assets and infrastructure in academia: A review on digital asset security. In *Complexities and Challenges for Securing Digital Assets and Infrastructure*, Pages 225–244.
- [11] Alshalaan, M., & Khan, N.A. (2025). Blockchain-enabled federated learning framework for secure and collaborative drug discovery: Integrating AI, molecular docking, and distributed ledger technology.
- [12] Alshreef, B.S., & Khan, N.A. (2025). LSTM neural networks in healthcare disruption forecasting: A comprehensive review and implementation framework. In *International Conference on Innovative Computing and Cutting-edge Technologies*, Springer Nature Switzerland Cham.
- [13] Alshreef, B.S., & Khan, N.A. (2026). Disruption forecasting: A comprehensive review and implementation framework. In *Advances in Smart Computing and Emerging Technologies: Proceedings of the Third International Conference on Innovative Computing and Cutting-edge Technologies (ICICCT 2025)*, Springer Nature.
- [14] Alsulami, M.H., Alotaibi, S., & Khan, N. (2021). Smart university model for Saudi Arabian universities. *Design Engineering*, Pages 162–181.
- [15] Alsulami, M.H., et al. (2021). Zigbee technology to provide elderly people with well-being at home. *International Journal of Sensors Wireless Communications and Control*, 11(9): 921–927.
- [16] Ayadi, W., Khalid, M.N., & Khan, N.A. (2026). Learning and fuzzy control. In *Advances in Smart Computing and Emerging Technologies: Proceedings of the Third International Conference on Innovative Computing and Cutting-edge Technologies (ICICCT 2025)*, Springer Nature.
- [17] Ayadi, W., et al. (2025). Enhancing anomaly detection in autonomous systems using deep learning and fuzzy control. In *International Conference on Innovative Computing and Cutting-edge Technologies*, Springer Nature Switzerland Cham.

- [18] Gangwar, S., et al. (2026). AI-driven signal processing for single-cell data analysis: A case study in cancer research. In *Sensor Signal Processing for Intelligent Systems*, Pages 171–194, Springer Nature Switzerland Cham.
- [19] Haidar, S.W., et al. (2026). A secure data collaboration framework using differential partial AI and federated learning: Architecture, simulation, and performance evaluation. *Procedia Computer Science*, 282: 1237–1242.
- [20] Haidar, S.W., et al. (2025). Deep learning driven automated system for speech emotion recognition. In *2025 IEEE 7th International Conference on Computing, Communication and Automation (ICCCA)*, IEEE.
- [21] Haidar, S.W., et al. (2025). Enhancing problem-solving skills through modelling-based AI and generative flow networks (GFN). In *2025 2nd Global AI Summit–International Conference on Artificial Intelligence and Emerging Technology (AI Summit)*, IEEE.
- [22] Hassan, M.A.A., Khan, N.A., & Nasim, M.A.M. (2017). Managing data replication in mobile ad-hoc network databases using content based energy optimization. *Mediterranean Journal of Basic and Applied Sciences*, 1(1): 142–154.
- [23] Khan, N.A. (2018). Cloud applications development and deployment: The future of cost-effective programming and a step ahead. *Middle East Journal of Applied Science & Technology*, 1(1): 30–36.
- [24] Khan, N.A. (2019). Wireless requirements and benefits in the academics domain. *Middle East Journal of Applied Science & Technology*, 2(3): 45–49.
- [25] Khan, N.A. (2019). *Basics of Ethical Hacking and Computer Security*. Paperback, 1 January 2019.
- [26] Khan, N.A. (2019). Security management protocols in cloud computation. *Middle East Journal of Applied Science & Technology*, 2(1): 16–23.
- [27] Khan, N.A. (2021). Measuring academics intentions to use a project management system (PMS): A case study of the College of Computing and Information Technology, Shaqra University. In *Trends in Future Informatics and Emerging Technologies*, Pages 58–69.
- [28] Khan, N.A. (2022). Development of an artificially intelligent advising system for Saudi medical transcription. *Development*, 6(3): 94796.
- [29] Khan, N.A. (2024). Development of intelligent pick and drop service manager for small cities. *Asian Journal of Basic Science & Research*, 6(3): 20–27.
- [30] Khan, N.A. (2024). Development of intelligent help system for small cities. *Asian Journal of Applied Science and Technology*, 8(3): 112–119.
- [31] Khan, N.A. (2025). Development of intelligent student information system. *Arabian Journal of Basic Science and Research*, 7(1).
- [32] Khan, N.A. (2025). An application of machine learning, big data, and IoT of enterprise architecture: Challenges. In *Quality Control–Artificial Intelligence, Big Data, and New Trends: Artificial Intelligence, Big Data, and New Trends*, 111.

- [33] Khan, N.A., & Ahamad, D. (2025). Living smart: AI-based urban assistance systems for sustainable wellbeing in small cities. *Middle East Journal of Applied Science & Technology*, 3(2): 89–99.
- [34] Khan, N.A., et al. (2021). Development of mubadarah system—an intelligent system for proposals at a university. In *2021 International Conference on Computational Intelligence and Knowledge Economy*, IEEE.
- [35] Khan, N.A., et al. (2025). Transformative impact of artificial intelligence on higher education: A comprehensive analysis of pedagogical innovation, institutional transformation, and future learning ecosystems. *Asian Journal of Applied Science and Technology*, 9(4): 57–76.
- [36] Khan, N.A., et al. (2021). Development of Medidrone: A drone based emergency service system for Saudi Arabian healthcare. In *2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, IEEE.
- [37] Khan, N.A., et al. (2026). A microservices-based optical mark recognition framework with advanced image processing and cloud-native architecture for scalable educational assessment. *Middle East Journal of Applied Science & Technology*, 9: 40–49.
- [38] Khan, N.A., et al. (2026). Design, development, and large-scale experimental evaluation of a multilingual, header-aware, course-centric intelligent attendance management and automated academic report generation framework for higher education institutions.
- [39] Khan, N.A., et al. (2026). User-centered development of an academic content website.
- [40] Khan, N.A., et al. (2025). Network intrusion management of web form spamming using blockchain. *Irish Interdisciplinary Journal of Science & Research*, 9(03).
- [41] Khan, N.A., et al. (2024). An IoMT enabled iterative artificial bee colony approach using federated learning for detection of heart disease. In *Solving with Bees: Transformative Applications of Artificial Bee Colony Algorithm*, Pages 103–116, Springer Nature Singapore.
- [42] Khan, N.A., et al. (2026). Revolutionizing drug design: The convergence of AI and quantum computing—A systematic review. In *Artificial Intelligence in Precision Drug Design*, Volume 2, Pages 317–326.
- [43] Khan, N.A., & Al Ghamdi, A.R. (2015). Cyber forensics and proposed techniques to overcome cyber threats for cyber security. *International Journal of Engineering and Management Research (IJEMR)*, 5(5): 187–191.
- [44] Khan, N.A., & Albatein, J. (2021). COVIBOT—An intelligent WhatsApp based advising bot for Covid-19. In *2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, IEEE.
- [45] Khan, N.A., & Alshalaan, M. (2025). AI-driven blockchain framework for digital transformation of academic accreditation process: A Saudi Arabian perspective.
- [46] Khan, N.A., & Alshalaan, M. (2026). Systematic review on machine learning applications in hardware security for vehicles. In *Software Engineering and Cybersecurity in Modern Electric Vehicles*, Pages 91–110.

- [47] Khan, N.A., et al. (2021). An empirical analysis on users' acceptance and usage of BYOD-technology for Saudi universities: A case study of Shaqra University. In 2021 International Conference on Technological Advancements and Innovations (ICTAI), IEEE.
- [48] Khan, N.A., Khan, A.R., & Rajeyyagari, S. (2025). Innovation in teaching and learning with the use of modern computational tools: A post Covid experience. *Middle East Journal of Applied Science & Technology*, 8(2): 74–82.
- [49] Khan, N.A., et al. (2026). Serverless absence management: A Google-native playbook for rapid, auditable university leaves workflows.
- [50] Khan, N.A., et al. (2026). A cloud-based multi-tier approval system for academic course registration: Design, implementation, and evaluations.
- [51] Khan, N.A., et al. (2020). Internet of Things (IoT) based educational data mining (EDM) system. *Journal of Mechanical and Computational Mathematics and Sciences*, 15(3): 271–284.
- [52] Khan, N.A., et al. (2019). Intrusion management to avoid web-form spamming in cloud based architectures. In 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), IEEE.
- [53] Khan, N.A., et al. (2025). A data-driven analysis of online learning expectations: Perspectives from Middle Eastern students and educators. In *International Conference on Innovative Computing and Cutting-edge Technologies*. Springer Nature Switzerland Cham.
- [54] Khan, N.A., et al. (2026). A data-driven analysis of online learning expectations: Perspectives. In *Advances in Smart Computing and Emerging Technologies: Proceedings of the Third International Conference on Innovative Computing and Cutting-edge Technologies (ICICCT 2025)*, Springer Nature.
- [55] Khan, N.A., Rajeyyagari, S., & Khan, A.R. (2025). Development of intelligent library services for university students. *Mediterranean Journal of Basic and Applied Sciences*, 9(1): 142–147.
- [56] Khan, N.A., Siddiqi, A.M.U., & Ahmad, M. (2021). Development of intelligent alumni management system for universities. *Asian Journal of Basic Science & Research*, 3(2): 51–60.
- [57] Khan, N.A., et al. (2019). Prevention of web-form spamming for cloud based applications: A proposed model. In 2019 Amity International Conference on Artificial Intelligence (AICAI), IEEE.
- [58] Kumar, D., & Dhir, S. (2021). Trends in Future Informatics and Emerging Technologies: Advanced Computing Techniques: Implementation, Informatics and Emerging Technologies, Volume 1.
- [59] Mahafdah, R.F., Al-Omari, O.M., & Khan, N.A. (2018). Learning modal adaptability to improve reading and writing skills of students.
- [60] Ren, M., et al. (2026). Explainable AI for structure-based drug design: Concepts and case studies. In *Artificial Intelligence in Precision Drug Design*, Volume 2, Pages 334–455, Elsevier.
- [61] Siddiqi, A.M.U., et al. (2026). Design and development. In *Advances in Smart Computing and Emerging Technologies: Proceedings of the Third International Conference on Innovative Computing and Cutting-edge Technologies (ICICCT 2025)*, Springer Nature.

[62] Siddiqi, A.M.U., et al. (2025). Design and development of a cost-effective software maintenance management model. In International Conference on Innovative Computing and Cutting-edge Technologies, Springer Nature Switzerland Cham.

[63] Zamani, A.S., Akhtar, M.M., & Khan, N.A. (2025). An application of machine learning, big data and IoT of enterprise architecture: Challenges, solutions and open issues. In Quality Control–Artificial Intelligence, Big Data, and New Trends, IntechOpen.

[64] Zamani, A.S., et al. (2026). Challenges in docking accuracy and reproducibility. In Recent Advances in Molecular Docking, Pages 435–446, Academic Press.