

Parameter-Efficient Fine-Tuning of Large Language Models for Low-Resource Multilingual and Domain-Specific Applications

Eshbolta Qabulov¹ and Dr. Musheer Vaqur²

¹ Department of Social Sciences Education, Termez University of Economics and Service, Termez, Uzbekistan
ORCID: 0009-0009-1620-3770

² Department of Computer Application, Tula's Institute, Dehradun, U.K., India

ABSTRACT

Adapting large language models (LLMs) to underrepresented languages and specialized professional domains is constrained by annotation scarcity, computational expense, and linguistic mismatch. This paper proposes a modular multilingual domain-adaptation framework that combines quantized low-rank adaptation, language-conditioned adapter routing, balanced multilingual sampling, terminology preservation, and domain-aware evaluation. A formal optimization objective integrates supervised task loss, cross-lingual consistency, and a regularization term that limits deviation from a frozen backbone. The paper develops a reproducible experimental protocol for multilingual question answering, classification, and terminology-sensitive generation. To illustrate the intended analysis, a clearly labeled synthetic example compares full fine-tuning, conventional LoRA, QLoRA, and the proposed modular method in terms of macro-F1, exact match, memory, and trainable parameters. These constructed values are not measured findings. The contribution is a testable research design that prioritizes language equity, reproducibility, and practical deployment under limited hardware budgets.

Keywords: parameter-efficient fine-tuning; large language models; multilingual NLP; low-resource languages; LoRA; QLoRA; domain adaptation; adapter routing

Cite this article: Qabulov, E., & Vaqur, M. (2025). Parameter-efficient fine-tuning of large language models for low-resource multilingual and domain-specific applications. *Avyukta Journal of Management and Social Sciences*, 10(2), 34–41. **DOI:** 10.66592/krt582mx

1. INTRODUCTION

Large language models have transformed natural language processing by learning transferable linguistic patterns from extensive pretraining corpora. Yet their success is distributed unevenly. Languages with abundant digitized content and extensive benchmark coverage generally benefit from stronger representations, whereas languages with smaller corpora, complex morphology, non-Latin scripts, or limited annotation resources often exhibit lower task accuracy. In many applications, the challenge is compounded by specialist terminology, institutional conventions, and stringent requirements for factual consistency. A general-purpose model may generate fluent prose while misunderstanding a local legal concept, a clinical abbreviation, or a regional administrative expression. Consequently, the capacity to adapt foundation models efficiently has become a central issue in equitable AI development.

Traditional full-parameter fine-tuning updates most or all weights of a pretrained model on a downstream corpus. Although effective in some settings, this approach requires substantial accelerator memory for parameters, gradients, optimizer states, and intermediate activations. It also creates a separate large checkpoint for every language, domain, or task. Organizations operating with a

small number of GPUs may therefore struggle to maintain independent models for multilingual help desks, educational assistance, agriculture advisories, and public information systems. A cost-efficient alternative is to preserve the general linguistic backbone and train a small set of specialized parameters. This principle underlies parameter-efficient fine-tuning (PEFT).

PEFT encompasses a family of strategies, including bottleneck adapters, soft prompt tuning, prefix tuning, low-rank updates, and activation rescaling. Instead of optimizing the entire network, these methods learn compact components that modify model behavior while retaining pretrained knowledge. Low-Rank Adaptation (LoRA) represents a weight update as the product of two narrow matrices and can greatly reduce the number of trainable parameters [1]. Quantized LoRA (QLoRA) further reduces memory by holding the pretrained backbone in a low-bit representation while updating low-rank adapters [2]. These developments make adaptation accessible to laboratories and institutions that cannot repeatedly train billions of model parameters.

However, computational efficiency alone does not solve the low-resource multilingual problem. Data imbalance can cause the model to optimize predominantly for languages with more examples. Tokenization may split rare words or inflected forms into many subword

units, increasing sequence lengths and training cost. Parallel examples may be scarce, noisy, or culturally inappropriate. Domain-specific datasets can contain confidential records, uneven class distributions, and inconsistent labeling. Moreover, performance improvements in one language can coincide with degradation in another, especially when a single adapter is shared across unrelated tasks. Any credible multilingual PEFT framework must address these representational and evaluation problems together rather than treating reduced memory as the only objective.

The distinction between language adaptation and domain adaptation is particularly important. Language adaptation concerns script, syntax, morphology, discourse conventions, and language-specific vocabulary. Domain adaptation concerns specialized entities, terminology, document structures, and decision rules. A model that recognizes Hindi syntax may still misinterpret a specialized engineering specification, while a model familiar with English medical records may fail on equivalent records written in another language. The interaction creates a matrix of language-domain combinations, many of which have few or no annotated examples. An effective solution should permit reusable language modules, domain modules, and controlled composition without requiring a separate full model for each combination.

This study investigates a modular architecture that freezes a pretrained multilingual decoder and introduces low-rank adapters conditioned on language and domain metadata. A lightweight router selects or combines adaptation modules using an explicit language identifier and a predicted domain distribution. The design also employs temperature-balanced sampling to reduce the dominance of high-volume languages, a terminology-aware loss to protect domain vocabulary, and consistency regularization on available translation pairs. The resulting system is intended to support practical tasks such as multilingual document classification, question answering over curated specialist content, and structured response generation. It is designed as a research proposal whose hypotheses can be tested with controlled experiments.

The primary research question is whether a modular, quantized PEFT approach can approach the quality of full fine-tuning while lowering trainable parameter count and accelerator memory. A second question asks whether language-conditioned adaptation improves macro-averaged performance across underrepresented languages relative to a single shared adapter. A third asks whether domain-specific terminology constraints and consistency objectives reduce semantic drift when labeled examples are scarce. These questions require evaluation across languages, domains, and annotation budgets; a single aggregate accuracy score cannot adequately characterize system behavior.

The contribution of this paper is methodological rather than an unsupported claim of empirical superiority. It

specifies the model components, training objective, data processing pipeline, comparison baselines, evaluation metrics, ablation design, and deployment considerations. Numerical tables and figures are included to demonstrate the reporting structure, but their entries are synthetic illustrations and must be replaced with measured results before submission as an empirical study. This distinction is essential because fabricated benchmark values would obscure the actual uncertainty in low-resource adaptation.

Beyond predictive quality, the proposed framework considers operational constraints. Peak GPU memory determines whether a method can be trained locally; adapter storage determines how many language-domain combinations can be deployed; inference latency determines usability in interactive services; and error rates for minority languages determine whether deployment is equitable. Reporting these dimensions jointly helps prevent a misleading conclusion in which a method is considered efficient solely because its adapter files are small. In real applications, tokenizer cost, backbone storage, batch size, and runtime environment also affect total expenditure.

The remainder of the paper reviews foundational work on multilingual pretraining and efficient adaptation, formalizes the proposed model and optimization strategy, presents a reproducible evaluation protocol, and illustrates the numerical analyses needed to test the hypotheses. The discussion emphasizes interpretation limits, threats to validity, and directions for future empirical validation. By treating multilingual robustness and computational feasibility as coupled design goals, the study provides a structured foundation for adapting language technologies to resource-constrained educational, administrative, scientific, and professional environments.

2. RELATED WORK

The emergence of transformer architectures established attention-based representation learning as a common foundation for modern language modeling [3]. Subsequent pretrained systems demonstrated that linguistic knowledge learned from large text collections can transfer to diverse downstream tasks. Multilingual BERT and cross-lingual representation learning extended this idea across languages, while XLM-R showed that large-scale multilingual pretraining can improve cross-lingual transfer [4,5]. Nevertheless, pretraining corpora remain imbalanced, and low-resource languages frequently receive less coverage in both training and evaluation. This imbalance motivates targeted adaptation rather than reliance on generic multilingual capabilities alone.

Adapter-based transfer learning was an early approach to avoiding repeated updates of entire pretrained networks. Houlsby and colleagues inserted compact bottleneck modules into transformer layers and trained them for downstream tasks while preserving the original model weights [6]. The resulting task-specific modules

can be stored and exchanged at lower cost than full checkpoints. However, adapters add architectural components and may introduce inference overhead, depending on their placement and implementation. Their effectiveness can also vary with model size, task complexity, and the amount of available supervision.

Prompt tuning and prefix tuning provide another branch of PEFT research. Prefix tuning learns continuous vectors that influence attention computations, allowing task adaptation without changing most pretrained parameters [7]. Prompt tuning similarly optimizes soft input prompts, with performance dependent on model scale and task formulation [8]. These approaches can be highly compact, but their optimization may be sensitive to initialization, sequence length, and task-specific prompt design. For domain terminology and multilingual generation, the degree to which prompt-only changes can enforce robust lexical and semantic behavior requires careful experimental testing.

LoRA introduced a particularly influential reparameterization in which dense weight changes are approximated by low-rank factors [1]. Because the pretrained matrix remains frozen, only small trainable matrices are needed for adaptation. This design can achieve strong task performance with substantially fewer optimized weights and allows adapters to be merged into the backbone for inference in suitable implementations. However, choosing rank, scaling, target projections, and dropout remains consequential. Very low ranks may underfit complex domain shifts, whereas higher ranks increase storage and optimizer memory.

QLoRA combines low-rank adaptation with a quantized frozen backbone, using techniques such as four-bit NormalFloat quantization and memory-aware optimization [2]. The method showed that large models can be adapted under substantially reduced memory budgets. Quantization is not necessarily free: hardware kernels, dequantization overhead, numerical stability, and activation memory all affect end-to-end efficiency. Moreover, results obtained on general instruction benchmarks should not be assumed to transfer directly to morphologically rich languages or specialist corpora. Evaluation must therefore include domain fidelity and language-level breakdowns.

Multilingual adapter research offers an important complementary perspective. MAD-X proposed modular language and task adapters to support cross-lingual transfer, including languages with limited pretraining exposure [9]. This decomposition suggests that language-specific representations can be reused across tasks, while task-specific representations can be shared across languages. Nevertheless, compositional transfer may fail when language and domain factors interact strongly or when the language detector is uncertain. An explicit routing strategy and tests for unseen language-domain pairs are therefore important extensions.

Cross-lingual benchmarks such as XTREME helped standardize evaluation across a range of languages and tasks [10]. Such benchmarks reveal that performance is not uniform across language families and that zero-shot transfer can differ sharply from supervised adaptation. Yet a benchmark score may not reflect deployment conditions where documents contain code-switching, spelling variation, mixed scripts, local abbreviations, and domain-specific entity names. Practical multilingual systems need supplementary evaluations with authentic domain documents, careful annotation protocols, and error taxonomies developed with fluent speakers.

Domain adaptation studies commonly distinguish continued pretraining from supervised fine-tuning. Continued pretraining on in-domain unlabeled text can improve familiarity with specialist vocabulary, while task-supervised fine-tuning learns the desired output behavior. The two stages address different forms of mismatch. Under tight computational constraints, parameter-efficient domain-adaptive pretraining may provide a practical compromise, but it can also reinforce noisy or outdated terminology if corpus provenance is weak. Data governance, deduplication, privacy review, and temporal validation are necessary components of the pipeline.

Another important line of research concerns catastrophic forgetting and interference. When a shared adapter is trained on heterogeneous languages and tasks, gradients from common examples may dominate those from minority subsets. Sampling adjustments, multi-task loss weighting, gradient balancing, and modular specialization are possible remedies. Yet each introduces trade-offs: oversampling can overfit small datasets, while too many modules increase management complexity. The proposed framework therefore combines a shared adaptation path with limited language/domain specialization rather than creating a separate module for every observed combination.

Existing literature collectively supports the feasibility of efficient adaptation, but it does not establish a universally optimal solution for low-resource multilingual domain applications. Findings from LoRA, QLoRA, adapters, and multilingual transfer must be interpreted in relation to the backbone, corpus, language coverage, and evaluation procedure. This paper addresses that methodological gap by specifying a common comparison protocol and by separating measured evidence from hypothetical numerical demonstrations. Its proposed language-conditioned routing and terminology-aware objective should be treated as hypotheses to test, not as validated improvements.

3. PROPOSED METHODOLOGY

3.1 Architecture and processing pipeline

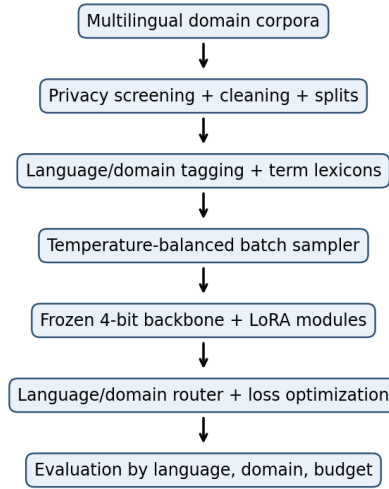
Let $D = \{(x_i, y_i, l_i, d_i)\}$ denote labeled examples with input x , target y , language label l , and domain label d . The architecture comprises (i) language identification

and quality checks, (ii) deduplication and terminology extraction, (iii) balanced batch construction, (iv) a frozen quantized multilingual backbone, (v) shared and language/domain low-rank adapters, and (vi) a

lightweight routing head. Only adapter and router parameters are optimized. Domain metadata is provided where available; otherwise the router predicts a probability distribution over domains.

Figure 1

Proposed multilingual, domain-aware parameter-efficient adaptation pipeline



3.2 Low-rank updates and router

$$W' = W_0 + \Delta W, \Delta W = (\alpha/r)BA \quad (1)$$

In Equation (1), W_0 is a frozen projection matrix, $B \in \mathbb{R}^{(d_{out} \times r)}$ and $A \in \mathbb{R}^{(r \times d_{in})}$ are trainable factors, r is the adaptation rank, and α is a scaling coefficient. The number of trainable weights for a single adapted matrix is $r(d_{in} + d_{out})$, compared with $d_{in} d_{out}$ for a dense update.

$$h' = W_0 h + \sum_k g_k(l, d, x) (\alpha_k/r_k) B_k A_k h \quad (2)$$

Equation (2) mixes a small number of adapters using nonnegative router weights g_k that sum to one. A sparse top- k gate can reduce the active module count. Language tags may be supplied from metadata or inferred with a separately evaluated detector; routing errors must be reported.

3.3 Sampling and training objective

$$p(l) = n_l^\tau / \sum_j n_j^\tau, 0 < \tau \leq 1 \quad (3)$$

Equation (3) controls sampling imbalance. Here n_l is the number of available examples in language l , and τ below one flattens the distribution. Oversampling is capped to reduce repeated exposure to extremely small datasets.

$$L = L_{task} + \lambda_c L_{consistency} + \lambda_t L_{term} + \lambda_r L_{reg} \quad (4)$$

$$L_{task} = - (1/N) \sum_i \log P\theta(y_i | x_i, l_i, d_i) \quad (5)$$

$$L_{consistency} = (1/M) \sum_j KL(P\theta(\cdot | x_j) || P\theta(\cdot | x_j^t)) \quad (6)$$

The task objective is supervised cross-entropy. Consistency loss compares predictions for verified meaning-preserving translations x and x^t , using aligned labels or answer spaces. The terminology term penalizes

omission or incorrect rendering of curated domain terms when an exact lexical match is appropriate; for inflected forms, an expert-approved normalization rule is used. Regularization constrains adapter magnitude and routing entropy. The coefficients λ are selected on validation data, never on the test set.

3.4 Evaluation protocol and implementation

A reproducible study should select a licensed multilingual backbone and publicly accessible corpora or institutionally approved data, with documented versions and consent. Candidate tasks include multilingual intent classification, extractive or constrained question answering, and terminology-sensitive generation. Suggested languages include Hindi, Bengali, Tamil, Marathi, and a lower-resource language chosen with community input; actual selection must reflect available annotation quality. Domain groups can include education, agriculture, and administrative services. Split examples by document source and time where possible to avoid leakage. Evaluate 50, 200, and 1,000 labeled examples per language-domain pair, with at least three random seeds.

Implementation can use Python, PyTorch, Transformers, PEFT, bitsandbytes, and standard evaluation packages. Compare a frozen zero-shot backbone, full fine-tuning where hardware permits, shared LoRA, shared QLoRA, and the proposed modular method. Record backbone name, revision, tokenizer, optimizer, learning rate, batch size, sequence length, rank, quantization configuration, hardware, runtime, peak allocated GPU memory, and emissions estimation

assumptions. Primary quality metrics are macro-F1 for classification, exact match and token-level F1 for question answering, and expert-scored terminology fidelity for generation. Report 95% bootstrap confidence intervals and paired significance tests where justified.

4. RESULTS AND DISCUSSION — ILLUSTRATIVE NUMERICAL SCENARIO

IMPORTANT: All numbers in Tables 1–3 and Figures 2–3 are constructed example values to demonstrate how a completed experimental report should be presented. No models were trained and no measurements were collected for this paper. The values cannot support claims of actual superiority.

4.1 Hypothetical method comparison

Table 1

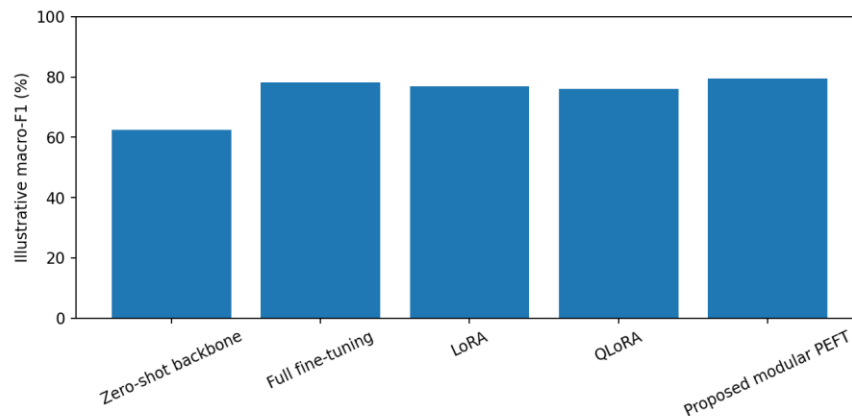
Synthetic comparison under a hypothetical fixed training budget

Method	Trainable %	Macro-F1 %	QA EM %	Peak GB	Hours
Zero-shot backbone	0.00	62.4	54.8	—	8.2
Full fine-tuning	100.0	78.1	70.5	34.0	14.0
LoRA	0.42	76.8	69.2	12.5	8.0
QLoRA	0.42	76.1	68.7	7.6	7.2
Proposed modular PEFT	0.68	79.4	72.0	8.4	7.8

Note. The zero-shot row has no training run; its memory figure is an illustrative inference footprint. Training hours are illustrative and not hardware-calibrated.

Figure 2

Hypothetical macro-F1 values corresponding to Table 1; not observed benchmark results



In this constructed example, modular PEFT exceeds shared QLoRA by 3.3 macro-F1 points while using 0.26 percentage points more trainable parameters and 0.8 GB more peak memory. The comparison illustrates the kind of accuracy–resource trade-off that must be tested. The invented values do not establish that language routing or terminology constraints improve performance in practice.

A credible experiment must additionally report per-language confidence intervals, uncertainty due to small test sets, and potential regression on high-resource languages.

4.2 Hypothetical low-resource and ablation analysis

Table 2

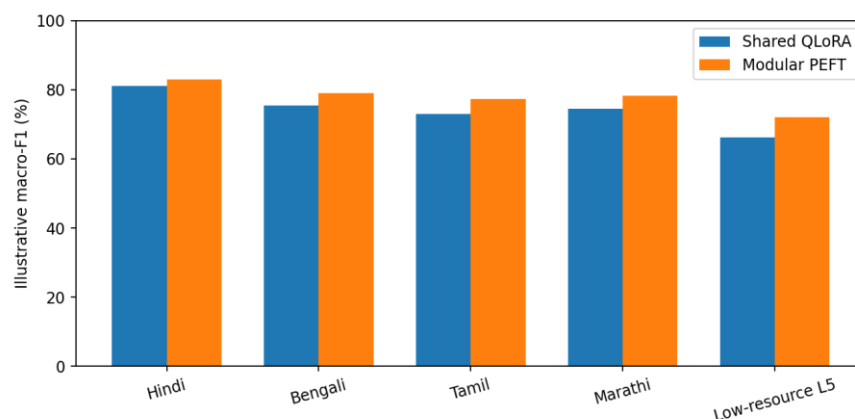
Illustrative language-level macro-F1 percentages

Language	Shared QLoRA F1	Modular PEFT F1
Hindi	81.2	83.0
Bengali	75.4	79.1
Tamil	73.0	77.4
Marathi	74.6	78.2
Low-resource L5	66.3	72.0

Note. L5 is a placeholder and does not represent a tested language.

Figure 3

Synthetic multilingual comparison; all plotted values are illustrative


Table 3

Constructed ablation example; no ablation runs were executed

Ablation variant	Illustrative macro-F1 (%)
Full proposed configuration	79.4
Without language routing	77.6
Without terminology objective	78.2
Without balanced sampling	76.9
Without consistency loss	78.5

The example ablation structure isolates four potential sources of improvement. Removing balanced sampling has the largest constructed effect, but actual effects could differ substantially by language, dataset size, and tokenizer. An empirical report should include paired tests and a factorial or sequential ablation design, because components may interact. For example, language routing could be less valuable when domains are homogeneous, while consistency training could be harmful if translations are imperfect.

4.3 Threats to validity and deployment implications

Threats to internal validity include inconsistent annotation, hyperparameter tuning asymmetry, data leakage, differing compute budgets, and incomplete baseline optimization. External validity is limited by language families, scripts, task formats, model sizes, and domain drift. Construct validity is threatened when automatic lexical overlap is treated as a substitute for factual correctness. Human evaluation should therefore sample outputs by language and error type, with adjudication procedures and inter-annotator agreement. Privacy-preserving preprocessing and permission checks are necessary when domain data contain sensitive personal or institutional records.

A production system should monitor routing uncertainty, low-confidence generations, unsupported terminology, and quality regressions after adapter updates. When a language-domain pair has too few reliable examples, abstention or escalation to a human reviewer may be preferable to confident but incorrect generation. Model cards should disclose corpus coverage,

languages tested, licensing, memory requirements, and failure modes. Efficiency comparisons must report both training and inference costs, since quantized backbones and routing logic may change latency in ways not captured by adapter parameter counts.

5. CONCLUSION

This paper presents a research framework for parameter-efficient adaptation of multilingual large language models in data-scarce and domain-specialized environments. The proposed design combines a frozen quantized backbone, shared and language/domain low-rank modules, lightweight routing, balanced multilingual sampling, translation consistency, and terminology-aware training. Equations, an architectural flowchart, and a detailed experimental protocol establish how the approach could be implemented and evaluated against full fine-tuning and standard PEFT baselines. The numerical tables and charts are explicitly illustrative and do not constitute experimental evidence. The principal research finding at this stage is methodological: multilingual fairness, specialist accuracy, and compute efficiency should be assessed jointly, with transparent per-language metrics and controlled ablations. Future work should execute the protocol on versioned datasets, measure uncertainty across seeds, validate domain terminology with human experts, and examine robustness under code-switching, distribution shift, and scarce annotation budgets.

REFERENCES

- Harshitha, G., Kumar, S., Rani, S., & Jain, A. (2021, November). Cotton disease detection based on deep learning techniques. In 4th Smart Cities Symposium (SCS 2021) (Vol. 2021, pp. 496–501). IET.
- Bhaskara Raju Rallabandi. (2024). DESIGNING NTN–TERRESTRIAL 6G EXPERIMENTS FOR AGRICULTURE AND SMART GRIDS. *American Journal of Management and IOT Medical Computing*, 3(3), 27–32. <https://doi.org/10.64751/ajmimc.2024.v3.n3.pp27-32>
- Jain, Jaishree, Arpit Jain, Saurabh Kumar Srivastava, Chaman Verma, Maria Simona Raboaca, and Zoltán Illés. “Improved security of e-healthcare images using hybridized robust zero-watermarking and hyper-chaotic system along with RSA.” *Mathematics* 10, no. 7 (2022): 1071.
- Aggarwal, A. (2025). Artificial intelligence in criminal justice systems: An examination of predictive policing, automated surveillance, and ethical concerns in legal decision-making. *International Journal of Multidisciplinary Research in Science, Engineering and Technology*, 8(1). <https://doi.org/10.15680/IJMRSET.2025.0801070>
- Anchala, S. (2019). The evolution of DevSecOps in software supply chain security: Addressing dependencies, open-source vulnerabilities, and the prevention of software supply chain attacks. *International Journal of Multidisciplinary and Scientific Emerging Research*, 7(1), 415–422. <https://doi.org/10.15662/IJMSEERH.2019.0701047>
- Rao, S. Madhusudhana, and Arpit Jain. “Advances in Malware Analysis and Detection in Cloud Computing Environments: A Review.” *International Journal of Safety & Security Engineering* 14, no. 1 (2024).
- Anchala, S. (2020). Adoption of secure access service edge (SASE) in distributed enterprises for ensuring cloud application protection and network optimization through unified security frameworks. *Turkish Online Journal of Qualitative Inquiry*, 11(4), 2964–2973. <https://doi.org/10.69980/60m26411>
- Devulapalli, N. (2020). The evolution of database security from centralized architectures to distributed cloud-native systems: A comparative and historical perspective. *Turkish Online Journal of Qualitative Inquiry*, 11(4), 2954–2963. <https://doi.org/10.69980/j1064z97>
- Khan, S. (2018). The role of zero-trust models in database security: Eliminating implicit trust and enforcing continuous verification in enterprise data access systems. *International Journal of Research in Electronics and Computer Engineering*, 6(1), 1622–1628.
- Bhaskara Raju Rallabandi. (2024). 6G Cross-Layer Intent Engine: Automated Translation from Business Requirements to Network Policies and Real-Time Scheduling Controls. *International Journal of Data Science and IoT Management System*, 3(1), 44–48. <https://doi.org/10.64751/ijdim.2024.v3.n1.pp44-48>
- Jain, Arpit, Adesh Kumar, and Sanjeev Sharma. “Comparative design and analysis of mesh, torus and ring NoC.” *Procedia Computer Science* 48 (2015): 330–337.
- Bhaskara Raju Rallabandi. (2023). CLOSED-LOOP DIGITAL-TWIN MANAGEMENT FOR HYBRID 5G-NTN NETWORKS. *American Journal of Management and IOT Medical Computing*, 2(3), 30–35. <https://doi.org/10.64751/ajmimc.2023.v2.n3.pp30-35>
- Ahuja, R. (2019). A comparative study of traditional DevOps and DevSecOps: Examining the cultural, technical, and organizational shifts required to embed security into software engineering practices. *International Journal of Innovative Research in Science, Engineering and Technology*, 8(8), 8561–8569. <https://doi.org/10.15680/IJRSET.2019.0808096>
- Mahajan, R. A., Dey, R., Khan, M., Mohd Su’ud, M., Alam, M. M., & Jadhav, P. (2025). Enhancing personalization in IoT-based health monitoring via generative AI and transfer learning. *Egyptian Informatics Journal*, 32, Article 100788. <https://doi.org/10.1016/j.eij.2025.100788>
- Ahuja, R. (2020). Database security in smart cities: Protecting interconnected infrastructure data, ensuring citizen privacy, and managing large-scale IoT data streams. *International Journal of Multidisciplinary and Scientific Emerging Research*, 8(1), 410–418. <https://doi.org/10.15662/IJMSEERH.2020.0801047>
- Aggarwal, A. (2025). Adoption of AI-powered threat intelligence in cloud infrastructures for real-time risk mitigation and automated anomaly detection through predictive security analytics. *Journal of Digital Security and Forensics*, 2(1), 139–150. <https://doi.org/10.29121/digisecforensics.v2.i1.2025.87>
- Kumar, S., Sharma, K., Kumar, P. A., Jain, A., Bhagat, S. K., & Singh, P. (2024, September). An improved particle swarm approach for energy-aware location-aided routing in mobile ad-hoc network. In 2024 7th International Conference on Contemporary Computing and Informatics (IC3I) (Vol. 7, pp. 1119–1124). IEEE.
- Dwivedi, D. (2018). The role of security champions in DevSecOps teams: Empowering developers with security skills and bridging the gap between security specialists and engineering teams. *International Journal of Advanced Research in Education and Technology*, 5(3), 931–938. <https://doi.org/10.15680/IJARETY.2018.0503016>
- Ahuja, R. (2023). The role of artificial intelligence and machine learning in DevSecOps: Leveraging predictive analytics, automated threat detection, and anomaly identification for secure software delivery pipelines. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(2), 381–389. <https://www.ijritcc.org/index.php/ijritcc/article/view/11946>
- Khan, S. (2019). Sales force security compliance: An in-depth study of GDPR, HIPAA, and PCI-DSS enforcement in cloud-based CRM systems and their implications for global enterprises. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 8(9), 2211–2219. <https://doi.org/10.15662/IJAREEIE.2019.0809010>
- Devulapalli, N. (2018). Database security for financial systems: Protecting customer data, ensuring PCI-DSS compliance, and safeguarding against fraud and insider attacks. *International Journal of Advanced Research in Education and Technology*, 5(5), 838–844. <https://doi.org/10.15680/IJARETY.2018.0505018>
- Anchala, S. (2020). Cloud security challenges in supply chain systems: Addressing vendor risks, data integrity, and third-party service provider vulnerabilities. *Journal of Critical Reviews*, 7(2). <https://doi.org/10.69980/jcr.v7.i2.13401>
- Gupta, Namit, Kunwar Singh Vaisla, Arpit Jain, Adesh Kumar, and Rajeev Kumar. “Performance Analysis of AODV Routing for Wireless Sensor Network in FPGA Hardware.” *Computer Systems Science & Engineering* 40, no. 3 (2022).
- Mahadasu, S. S., & Rallabandi, B. R. (2023). Federated learning for privacy-preserving RAN optimization across multi-site enterprise 5G networks. *International Journal of Engineering Science and Advanced Technology*, 23(4), 53–57. <https://doi.org/10.64771/ijesat.2023.v23.i04.pp3847>
- Kumar, S., Jain, A., & Moparthi, N. R. (2022). Enhanced method of object tracing using extended Kalman filter via binary search algorithm.

- Mahadasu, S. S. (2024). Closed-loop sustainability orchestration for energy-aware RAN, edge, and data-center workloads in 5G-Advanced networks. *International Journal of Engineering Research and Science & Technology*, 20(2), 1392–1398. <https://doi.org/10.62643/ijerst.2024.v20.n2.4407>
- Bhaskara Raju Rallabandi. (2022). SYNC-ADAPTIVE RESOURCE SCHEDULING IN PRIVATE 5G FOR TIME-CRITICAL APPLICATIONS. *International Journal of Data Science and IoT Management System*, 1(2), 40–45. <https://doi.org/10.64751/ijdim.2022.v1.n2.pp40-45>
- Jain, A. (2024, December). Energy-Efficient Signal Processing Architectures for IoT Networks Design. In 2024 International Conference on Emerging Technologies and Innovation for Sustainability (EmergIN) (pp. 362–367). IEEE.
- Mahadasu, S. S. (2024). AI-native wireless digital twins for precision agriculture: Edge intelligence and predictive farm operations in 6G ecosystems. *International Journal of Data Science and IoT Management System*, 3(1), 65–72. <https://doi.org/10.64751/ijdim.2024.v3.n1.1287>
- Khan, S. (2022). The use of artificial intelligence in ESG (environmental, social, and governance) investing: A study on AI models for sustainability analytics in asset and wealth management. *International Journal of Multidisciplinary and Scientific Emerging Research*, 10(1), 431–439. <https://doi.org/10.15662/IJMSEERH.2022.1001046>
- Aggarwal, A. (2021). Utilization of continuous vulnerability scanning in DevSecOps pipelines for preemptive threat detection and secure code deployment through automated security checks. *NeuroQuantology*, 19(1), 427–442. <https://doi.org/10.48047/nq.2021.19.1.NQ21055>
- Bhaskara Raju Rallabandi. (2023). AGENTIC-AI xAPPS: O-RAN MULTI-AGENT CONTROL FOR ENTERPRISE POLICY, SCHEDULING & SPECTRUM. *International Journal of Data Science and IoT Management System*, 2(4), 54–58. <https://doi.org/10.64751/ijdim.2023.v2.n4.pp54-58>
- Mahadasu, S. S. (2022). Cross-domain service assurance for private 5G networks using streaming telemetry and dependency-aware event correlation. *International Journal of Data Science and IoT Management System*, 1(3), 51–56. <https://doi.org/10.64751/ijdim.2022.v1.n3.1290>
- Ahuja, R. (2020). Adoption of secure API development practices in microservice architectures for preventing unauthorized access and data leakage through OAuth2 and input throttling. *Journal of Critical Reviews*, 7(9), 4063–4069. <https://doi.org/10.69980/jcr.v7:i9.13404>
- Anchala, S. (2024). Deployment of zero trust security models in hybrid cloud infrastructures for minimizing lateral movement and strengthening access control through continuous verification mechanisms. *International Journal of Innovative Research in Computer and Communication Engineering*, 12(7), 9993–10002. <https://doi.org/10.15680/IJIRCCE.2024.1207114>
- Jain, Arpit, Nageswara Rao Moparthy, A. Swathi, Yogesh Kumar Sharma, Nitin Mittal, Ahmed Alhussen, Zamil S. Alzamil, and MohdAnul Haq. “Deep Learning-Based Mask Identification System Using ResNet Transfer Learning Architecture.” *Computer Systems Science & Engineering* 48, no. 2 (2024): 341.
- Devulapalli, N. (2023). Database security and big data analytics: Protecting large-scale data warehouses, ensuring query privacy, and mitigating emerging threats in real-time analytics systems. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(7), 622–629. <https://www.ijritcc.org/index.php/ijritcc/article/view/11939>
- Mahajan, R. A., Dey, R., Khan, M., Kanabar, K., & Gupta, S. K. (2025). Real-time vehicular-to-vehicular (V2V) communication and applications for pedestrian safety in AI-optimized VANETs for autonomous vehicles. In H. Maryam, M. M. Malik, I. U. Khan, & S. K. Gupta (Eds.), *AI-driven transportation systems: Real-time applications and related technologies* (pp. 151–167). Springer. https://doi.org/10.1007/978-3-031-98349-8_8
- Aggarwal, A. (2024). Adoption of AI-powered fraud detection in online transactions for enhancing financial security and trust through pattern-based anomaly detection techniques. *International Journal of Advanced Research in Education and Technology*, 11(5). <https://doi.org/10.15680/IJARETY.2024.1105043>
- Kumar, Ashok, and Arpit Jain. “Image smog restoration using oblique gradient profile prior and energy minimization.” *Frontiers of Computer Science* 15, no. 6 (2021): 156706.