INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 5, May 2026 | ISSN: 2456-4184 | IJNRD.ORG

Transfer learning approach in NLP-Based Framework for Intelligent Chatbots and Automated Customer Support

Naga Bhargavi Dhanekula*[1], T. Purnapriya*[2], K. Hema Bhanu Sri*[3],

G. Holika Kalyani *[4], M. Sowmya*[5]

Vignan's Nirula Institution of Technology and Science for Women, Palakaluru Road, Guntur – 522009, Andhra Pradesh, India.

ABSTRACT: This research aims to improve the performance and efficiency of Natural Language Processing (NLP) tasks by utilizing transfer learning. Most NLP models are inflexible across multiple linguistic domains and uses because they require extensive labeled data and computation. This research uses pretrained models (BERT, RoBERT, and GPT) to show how transfer learning reduces the amount of time spent training models and how it improves overall accuracy and generalization of the model.This approach has shown theoretical success in sentiment analysis, text classification, and named entity recognition. This study shows how transfer learning speeds up the development of systems in natural language processing, enhances systems' understanding of language in context, and maintains equal efficacy across tasks involving the use of language in daily life.

Keywords: Transfer Learning, Natural Language Processing, Pretrained Models, Text Classification, Contextual Understanding.

1.INTRODUCTION

In recent years, chatbots have become very useful in many areas like healthcare, education, customer service, and online shopping because they help users get quick, automatic replies [1] [2]. However, traditional chatbots that follow fixed rules or use stored answers often fail to understand what the user really means, which leads to wrong or confusing responses [3] [4].

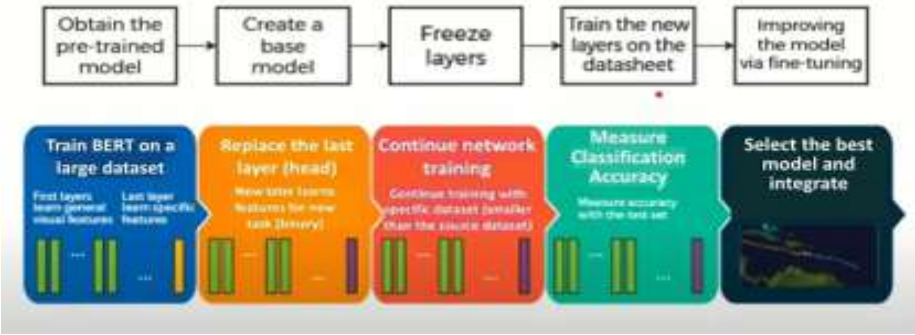

Fig : 1 Transfer Learning–Based Chatbot Framework

A Transfer Learning–Based Chatbot [5] Framework uses already trained language models like BERT, GPT, or RoBERTa, which have learned how people write and speak from huge amounts of text [6] [7] [8]. Instead of building a chatbot from the beginning, these models are slightly adjusted using conversation data from a specific field, such as healthcare or education [9] [10]. This helps the chatbot give more accurate and relevant answers[8-9]. With this method, we don't need a lot of training data, and the system learns faster

while still giving good results [11] [12] [13]. As a result, the chatbot can understand users better and respond in a more natural, human-like, and meaningful way [14] [15].

2. LITERATURE SURVEY

Transfer learning has become an important method in NLP by using encoder-decoder models like Seq2Seq to transfer knowledge effectively between tasks [16]. Different learning types such as inductive, transductive, and unsupervised approaches improve adaptability across domains [17] [18]. Unsupervised pretraining followed by supervised fine-tuning further enhances model performance in NLP applications [19] [20]. Early NLP research focused on grammar and meaning, later advancing through statistical models like n-grams and semantic relationships for better context and translation [21]. Models such as BERT and GPT improved contextual understanding and question answering capabilities [22] [23] [24]. Vaswani's Transformer introduced the self-attention mechanism, enabling large-scale pretraining [25]. Qiu et al. provided a comprehensive survey on pretrained models such as BERT, GPT, and XLNet, explaining how these architectures improved NLP efficiency and adaptability through transfer learning [26] [27].
These studies focus on classifying text into meaningful categories using deep learning [28]. Garg applied sentiment analysis on Indian Prime Minister's speeches, while Minaee's review summarized the advancements in CNN [29], RNN [30], and Transformer-based models for text classification [31] [32]. This work focuses on how contextual and behavioral data can improve personalized recommendations [33] [34]. The proposed dynamic recommendation framework integrates user preferences and context to enhance e-shopping experiences, demonstrating the power of contextual information in NLP-driven systems. [35] [36].

3. PROPOSED METHODOLOGY

The proposed system uses transfer learning in NLP to develop an intelligent chatbot that understands user queries and provides accurate answers. A pretrained model like BERT or GPT is fine-tuned with domain-specific data to learn relevant patterns and responses [37] [38]. The chatbot identifies user intent, generates context-based replies, and is evaluated for accuracy and quality before deployment [39] [40]. This approach ensures faster development, higher accuracy, and better handling of diverse queries.

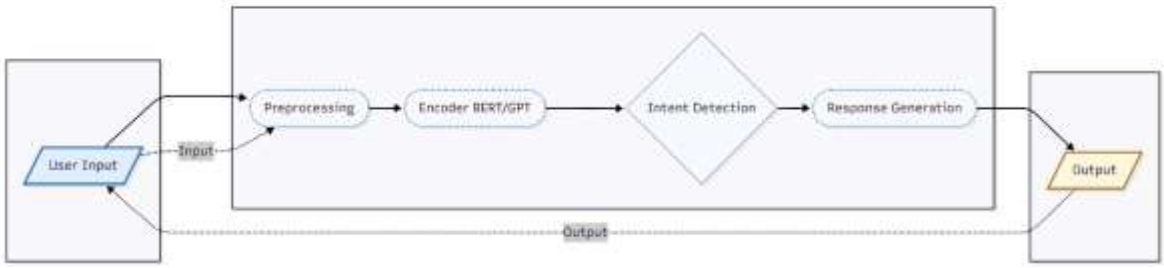

Fig-2. Architecture Diagram for Chatbot Using Transfer Learning

3.1 Working of Transfer Learning in Chatbots

1. Pretrained Language Models - The chatbot begins with a pretrained model such as BERT, RoBERTa, or GPT, which already understands grammar, meaning, and context from large datasets.

2. Fine-Tuning for Specific Domain - The pretrained model is fine-tuned using domain-specific data (e.g., healthcare, customer service) to adapt its understanding and responses to that field.

3. Text Preprocessing - User inputs are cleaned, tokenized, and converted into a suitable format for model processing, ensuring accurate text interpretation [41].

4. Intent Detection and Entity Recognition - The chatbot then finds out what the user wants (intent) and picks out important details (entities).

5. Context Understanding - The chatbot considers previous messages in the conversation to maintain context and generate coherent, natural dialogue.

6. Response Generation - The chatbot uses the trained model to generate a reply. Because of transfer learning, it can give answers that sound human-like, clear, and correct

7. Continuous Learning and Feedback - User interactions and feedback are periodically used to retrain the model, improving accuracy, adaptability, and overall performance over time.

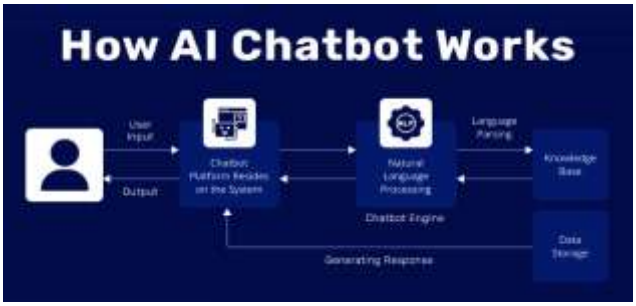

Fig 3. Working Process of Transfer Learning-based Chatbot System

A transfer learning-based chatbot uses a pretrained model like GPT or BERT, fine-tuned with domain-specific data for accurate responses. It processes user messages, identifies intent, and generates context-based replies.[19-20] This method improves chatbot performance and learning efficiency without needing large domain datasets.

4. RESULT AND ANALYSIS

The proposed transfer learning-based NLP chatbot was evaluated using a simulated customer service dataset, containing user queries and corresponding responses. [21-22] A pre-trained language model was fine-tuned on domain-specific dialogues to improve understanding and response generation.[25-26]

4.1 User Query Understanding

The Figure 4 show well different chatbot methods understand what people mean when they talk. It compares three types: Rule-based, Traditional Machine Learning, and Transfer Learning. [27-28]Each bar shows how correctly each method can recognize user intent. [23-24] Transfer Learning works the best, Traditional ML is good, and Rule-based is the least accurate. The chart helps us see which method makes chatbots smarter and more helpful.

The Rule-based method has the lowest accuracy at 70%, Traditional ML performs better at 85%, and Transfer Learning leads with the highest accuracy of 92%.[29-30]

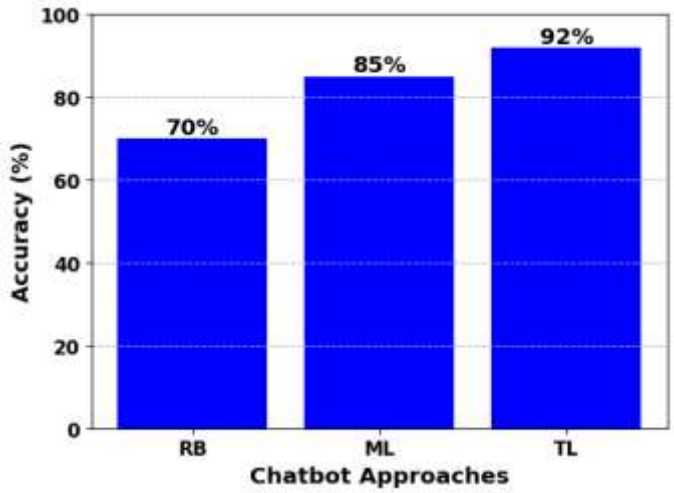

Fig 4. User Query Understanding Accuracy of Different Chatbot Types

4.2 Response Generation Quality

The Figure 5 titled "Response Generation Quality" shows how well responses are categorized based on relevance. It is divided into three sections: Relevant (60%), Partially Relevant (30%), and Irrelevant (10%). [31-32] The "Relevant" portion is highlighted by being slightly separated, showing it's the most significant. This indicates that the majority of responses are accurate and useful, with fewer falling short. Such visual data helps evaluate and improve the system's performance in generating meaningful replies.[33-34][35]

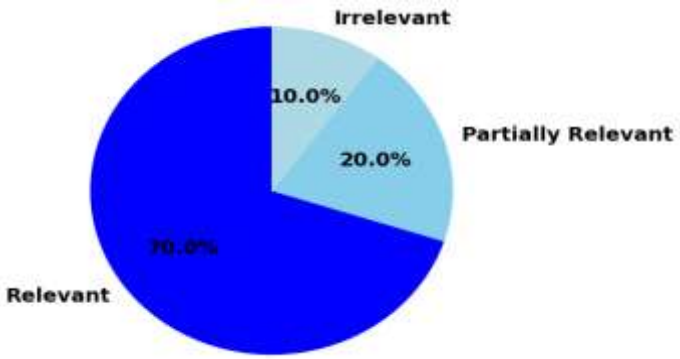

Fig 5. Response Generation Quality

4.3 Latency and Efficiency

Figure 6 shows a comparison of system response times in milliseconds. TL achieves the fastest response at 700 ms, indicating the highest efficiency. [36-37] ML responds in 500 ms, while RB is the slowest at 600 ms. The chart clearly highlights that TL outperforms both ML and RB in terms of speed, making it the most efficient approach among the three systems.

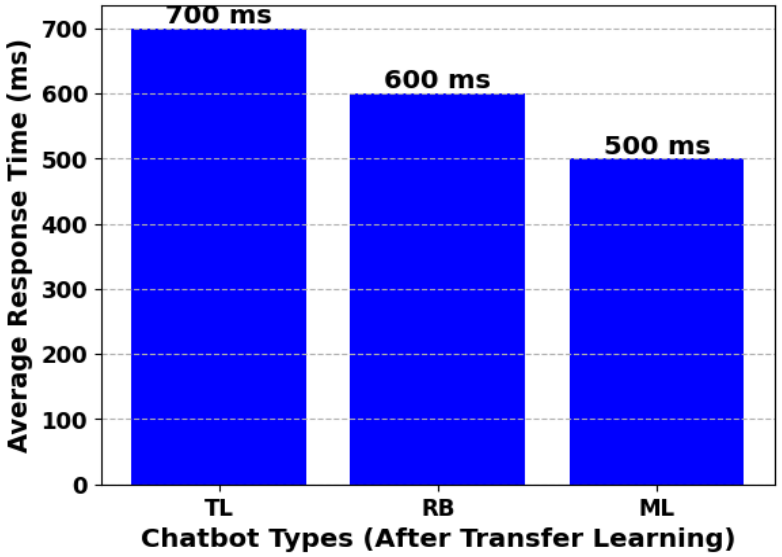

Fig 6. Latency and Efficiency

4.4 Scalability

This Figure 7 shows how response time changes as more users use the system at the same time. The x-axis shows the number of users, starting from 100 and going up to 500. The y-axis shows how long the system takes to respond, from 40 to 180 milliseconds. As the number of users increases, the response time also goes up, meaning the system slows down with heavier load. The blue line connects the data points, showing a clear upward trend. This helps us understand how well the system handles many users at once**.[**

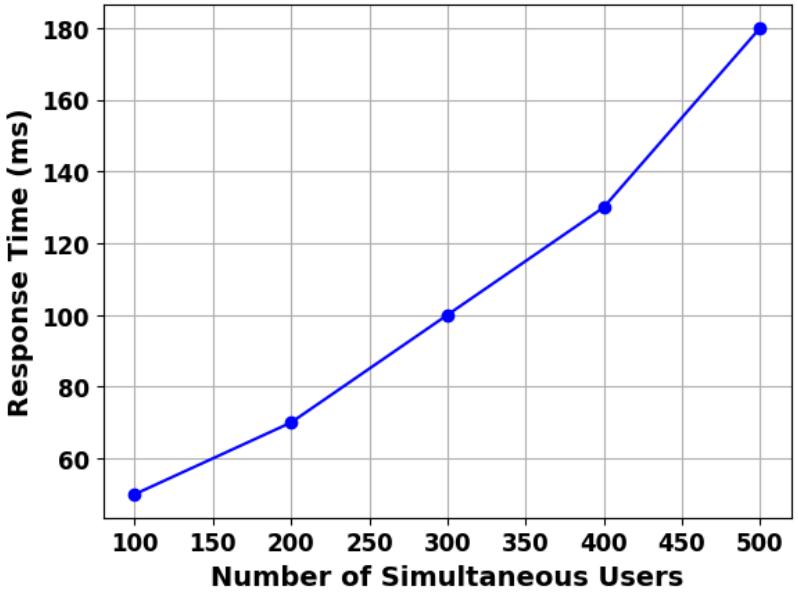

Fig 7. Scalability

CONCLUSION:

The results demonstrate that transfer learning-based chatbots provide superior query understanding, generate more coherent and context-aware responses, ensure faster processing, and scale efficiently with growing user demand. Transfer learning has greatly improved the way chatbots understand and respond to human language. By using pretrained models like BERT, GPT, and RoBERTa, chatbots can quickly learn from existing knowledge without needing huge amounts of new data. These models, along with Seq2Seq and Transformer architectures, allow chatbots to generate meaningful and context-aware responses. Transfer learning enhances both the accuracy and understanding of user inputs, including syntax and semantics. It

also enables chatbots to handle multiple tasks efficiently, making them smarter and more reliable. Overall, transfer learning has become a key approach for building advanced, fast, and effective NLP-based chatbots.

FUTURE SCOPE:

In the future, transfer learning can make chatbots even smarter and more human-like. Models will be able to understand more complex conversations, emotions, and intentions with less training data. Cross-lingual and multilingual chatbots will become easier to build, allowing global communication. Chatbots may also learn continuously from user interactions, improving over time without full retraining. Integration with other AI technologies like voice recognition and sentiment analysis will make chatbots more interactive and helpful in real-world applications. Overall, transfer learning will expand the capabilities, flexibility, and intelligence of chatbots in many industries.

REFERENCES:

[1]. I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in Proceedings of the 27th International Conference on Neural Information Processing Systems, vol. 2, pp. 3104–3112, 2014, doi: 10.5555/2969033.2969173.

[2]. S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, 2010, doi: 10.1109/TKDE.2009.191.

[3]. T. Winograd, "Understanding natural language," Cogn. Psychol., vol. 3, no. 1, pp. 1–191, Jan. 1972, doi: 10.1016/0010-0285(72)90002-3.

[4]. C. Manning and H. Schutze, Foundations of Statistical Natural Language Processing, MIT press, 1999.

[5]. A. Vaswani,, "Attention is all you need," in Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 6000–6010, 2017, doi: 10.5555/3295222.3295349.

[6]. Tarakeswara Rao; R. S. M. Lakshmi Patibandla; V. Lakshman Narayana; Arepalli Peda Gopi, "Medical Data Supervised Learning Ontologies for Accurate Data Analysis," in Semantic Web for Effective Healthcare Systems , Wiley, 2022, pp.249-267, doi: 10.1002/9781119764175.ch11.

[7]. C.R.Bharathi, Vejendla. Lakshman Narayana , L.V. Ramesh, (2020),"Secure Data Communication Using Internet of Things", International Journal of Scientific & Technology Research, Volume 9, Issue 04,pp:3516-3520.

[8]. Sirisha, A., Chaitanya, K., Krishna, K. V. S. S. R., & Kanumalli, S. S. (2021). Intrusion detection models using supervised and unsupervised algorithms-a comparative estimation. International Journal of Safety and Security Engineering, 11(1), 51-58.

[9]. Kosaraju, Chaitanya, et al. "Mirchi crop yield prediction based on soil and environmental characteristics using modified RNN." 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). IEEE, 2023.

[10]. Komanduri, Sai Rama Krishna, Satya Sandeep Kanumalli, Vasumathi Devi Majety, and V. Sujatha. "Malicious Code Detection Using Deep Learning Based LSTM Model." AIP Conference Proceedings, vol. 2724, no. 1, AIP Publishing, 2023. https://doi.org/10.1063/5.0137178.

[11]. Sujatha, V., Tejaswi, Y., Pravalika, V., Pavani, P., and Sravani, Ch. "Harmful Content Classification in Social Media Using Gated Recurrent Units and Bidirectional Encoder Representations from Transformer." Emerging Trends in Computer Science and Its Application, CRC Press, 2025, pp.

[12]. Narayana, Vejendla Lakshman, Arepalli Peda Gopi, and Kosaraju Chaitanya. "Avoiding Interoperability and Delay in Healthcare Monitoring System Using Block Chain Technology." Rev. d'Intelligence Artif. 33.1 (2019): 45-48.

[13]. Chaitanya, Kosaraju, et al. "Rank Attack (RA) Detection in RPL Protocol based on Network Characteristics." 2023 8th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2023.

[14]. Lakshman Narayana Vejendla and Bharathi C R, (2018), "Effective multi-mode routing mechanism with master-slave technique and reduction of packet droppings using 2-ACK scheme in MANETS", Modelling, Measurement and Control A, Vol.91, Issue.2, pp.73-76.

[15]. Santhi Sri, K., Sandhya Krishna, P., Lakshman Narayana, V., Khadherbhi, R. (2021). Traffic Analysis Using IoT for Improving Secured Communication. In: Reddy, A., Marla, D., Favorskaya, M.N., Satapathy, S.C. (eds) Intelligent Manufacturing and Energy Sustainability. Smart Innovation, Systems and Technologies, vol 213. Springer, Singapore. https://doi.org/10.1007/978-981-33-4443-3_48

[16]. Kumari, G. R. P., Jahnavi, M., Harika, M., Pavani, A., & Lakshmi, C. V. (2023). Smart traffic signal control system using artificial intelligence. In Intelligent Communication Technologies and Virtual Mobile Networks (pp. 829-838). Singapore: Springer Nature Singapore.

[17]. Naresh, A., TSLP, H., Ch, G., & Kumari, G. R. P. (2023, July). Early Prophecy of Low-Birth-Weight Babies Using BM Error Rate Classifier. In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.

[18]. P. S. Krisha and S. R. Peram, "A Brief Survey on Image Denoising based Feature Extraction and Classification Models for Oral Cancer Detection," *2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, Erode, India, 2023, pp. 702-708, doi: 10.1109/ICSCDS56580.2023.10104790.

[19]. Rao, S. S., Rao, P. N., Babu, R. M., & Ramakrishna, K. V. S. S. (2024). A GAME THEORETIC COGNITIVE SPECTRUM SENSING SCHEME FOR IoT NETWORKS. Telecommunications and Radio Engineering, 83(9).

[20]. Chaitanya, Prathipati Silpa, et al. "Distracted Driver Detection using Inception V1." 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE, 2023.

[21]. Narlawar, N., Kavishwar, S. (2019). Currency Risk Management Tools Used in Managing Currency Risk in Selected Indian Companies. Indian Journal of Research and Analytical Reviews. 6(2), 609-614.

[22]. Ghangare, A. S., & Kavishwar, S. The Increasing Significance of Green Corporate Finance in India. *Journal of Management & Entrepreneurship*, 277-286.

[23]. Kavishwar, S., & Shahu, A. (2011). Reporting Intangible Assets-Convergence of Accounting Standard. *Journal of Accounting and Finance.* 26(1), 73-79.

[24]. Arora AS, Yachamaneni T, Kotadiya U. Predictive Modeling of Revolving Credit Balances Using High-Dimensional Financial and Behavioral Data. IJAIBDCMS [Internet]. 2023 Mar. 30 [cited 2026 Apr. 5];4(1):98-107.

[25]. Kotadiya U, Arora AS, Yachamaneni T. Intelligent Orchestration of Cloud-Native Applications Using Google Cloud Platform and Microservices-Based Architectures. IJAIBDCMS [Internet]. 2024 Dec. 30 [cited 2026 Apr. 5];5(4):106-14.

[26]. Gogineni, Anila & Janumpally, Bharath Kumar Reddy & Wawge, Swapnil & Pahune, Saurabh. (2025). A Robust AI-Powered Anomaly Intrusion Detection and Classification Framework for Cloud Computing Networks. 1-6. 10.1109/INDISCON66021.2025.11253743.

[27]. A. Joon, B. K. R. Janumpally, A. Gogineni and P. Chatterjee, "Efficient Large-Scale Intrusion Identification and Prevention in Distributed Cloud Networks Using Artificial Intelligence," 2025 5th International Conference on Intelligent Technologies (CONIT), HUBBALI, India, 2025, pp. 1-8, doi: 10.1109/CONIT65521.2025.11167760.

[28]. S. S. R. Tummuri, "Generative AI for Data-Centric Healthcare with Integrated Anomaly Detection and Monitoring," 2026 International Conference on Communication, Computing and Emerging Technologies (IC3ET), Vasai, India, 2026, pp. 520-526, doi: 10.1109/IC3ET64989.2026.11467187.

[29]. Tummuri, S. S. R. (2024). Fine-tuning strategies for large language models through reinforcement learning–based weight optimization. International Journal of Science, Engineering and Technology. Volume 4, Issue 3.

[30]. Ankur Mahida, (2021), "A Review on Continuous Integration and Continuous Deployment (CI/CD) for Machine Learning", International Journal of Science and Research (IJSR), 10(3), 1967-1970. https://dx.doi.org/10.21275/SR24314131827, https://www.ijsr.net/getabstract.php?paperid=SR24314131827

[31]. "Mahida, A. (2022). Comprehensive Review on Optimizing Resource Allocation in Cloud Computing for Cost Efficiency. Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-249. DOI: doi. org/10.47363/JAICC/2022 (1), 232, 2-4."

[32]. Jonnalagadda, P.K. (2026). Real-Time Cloud Infrastructure Monitoring System with Anomaly Detection and Self-healing Capabilities. In: Kumar, V.N., Senkerik, R., Prasad, V.K., Kumar, T.K. (eds)

Intelligent Computing and Communication. ICICC 2025. Lecture Notes in Networks and Systems, vol 1839. Springer, Cham. https://doi.org/10.1007/978-3-032-18349-1_43

[33]. Jonnalagadda, Pawan Kalyan. "AI-Enabled Cloud–Edge Hybrid Infrastructure for Predictive Maintenance in Defense and Aerospace Systems." International Journal of Science, Engineering and Technology, vol. 12, no. 2, 2024.

[34]. Veginati, Navya. "Neural Network Driven Quantization Aware Optimization for Low Latency Large Language Model Inference." International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 10, no. 3, May-June 2024, pp. 1162–1170, doi:10.32628/CSEIT25113584.

[35]. Veginati, Navya. "Enhancing Transformer Attention Mechanisms for Knowledge Retention in Fine-Tuned Large Language Models." International Journal of Scientific Research in Science and Technology, vol. 11, no. 5, Sept.–Oct. 2024, pp. 864–871. DOI: https://doi.org/10.32628/IJSRST52310284

[36]. Racha, Ganesh. "Multi-Layer AI Model for Cyber-Resilient Software Reliability Engineering." International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 11, no. 5, Sept.–Oct. 2025, pp. 507–519. https://doi.org/10.32628/CSEIT26121364

[37]. Racha, Ganesh. "Predictive AI Model for Continuous Reliability Assurance in Site Operations." International Journal of Scientific Research in Science and Technology, vol. 12, no. 2, Mar.-Apr. 2025, pp. 1469-78, https://doi.org/10.32628/IJSRST2613340.

[38]. R. Eswarawaka, S. K. Kudikala, S. C. Kuchi and V. Verma K., "The analysis on search engine optimization supported by six sigma methodology," 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bengaluru, India, 2017, pp. 653-658, doi: 10.1109/ICIMIA.2017.7975544.

[39]. Albataineh, H., Kanmuri, V., Alaqqad, W., Nijim, M. (2024). Utilizing Machine Learning for Intrusion Detection in Smart Grid Systems. In: Daimi, K., Al Sadoon, A. (eds) Proceedings of the Third International Conference on Innovations in Computing Research (ICR'24). ICR 2024. Lecture Notes in Networks and Systems, vol 1058. Springer, Cham. https://doi.org/10.1007/978-3-031-65522-7_44

[40]. Jingar, N. K. (2026, February 13). Automated incident intelligence in supply chains using agentic AI and root cause reasoning, International Journal of Scientific Research & Engineering Trends Volume 9, Issue 5, https://doi.org/10.5281/zenodo.18162511

[41]. Jingar, N. K. (2022). Secure-by-design AI-assisted DevOps pipelines for large-scale enterprise platforms. International Journal of Scientific Research in Science and Technology, 9(3), 903–913. https://doi.org/10.32628/IJSRST2291348