Beyond the Human Signature: Transformer-Based Detection of AI-Generated Phishing Emails and Social Engineering Attacks

Paul Anderson*

Maricopa County Community College District

Abstract

The emergence of large language models capable of producing fluent, grammatically flawless prose has quietly dismantled one of the oldest heuristics in email security: the assumption that clumsy language betrays a phishing attempt. This paper synthesizes recent scholarship on transformer-based detection of phishing emails and social engineering messages, with particular attention to the harder problem of identifying content generated or substantially rewritten by large language models. Adopting a theoretical and integrative research design rather than a new empirical experiment, the study draws on peer-reviewed and reputable preprint literature published mainly between 2022 and 2026 to trace how encoder-based transformers such as BERT and RoBERTa, distilled variants such as DistilBERT, and emerging large language model detectors have been applied to this problem, and to identify where their performance claims break down under adversarial or out-of-distribution conditions. The synthesis finds that transformer architectures deliver consistently high accuracy on established benchmark corpora, frequently exceeding ninety-eight percent, yet this performance depends heavily on corpus composition and degrades measurably when models encounter phishing campaigns authored by generative systems rather than by the human attackers represented in legacy datasets. The paper develops a conceptual comparison of model families, spanning encoder-only transformers, distilled architectures, hybrid neural designs, large language model detectors, and stylometric or perplexity-based approaches, mapping the strengths and failure modes associated with each. The discussion situates these findings within a broader argument that detection can no longer rest on linguistic anomaly alone and must instead incorporate behavioral, contextual, and provenance signals alongside textual analysis. The paper closes with recommendations for adversarially robust benchmarking and human-in-the-loop deployment.

Journal of Science, Technology and Social Transformation (2025)

Introduction

Phishing remains among the most persistent vectors of cyber intrusion precisely because it targets human judgment rather than technical infrastructure, and for two decades the discipline of automated phishing detection has relied substantially on the premise that deceptive messages carry detectable linguistic residue: awkward phrasing, inconsistent register, or telltale grammatical error. That premise is now under direct strain. Generative language models can produce social engineering content that is fluent, contextually tailored, and free of the surface errors that earlier classifiers were trained to notice, a development that several recent studies describe as having already been demonstrated against commercial large language models tasked with generating convincing phishing text. The practical consequence is that detection systems calibrated against corpora of human-authored phishing may systematically underperform against a new and rapidly growing class of machine-authored social engineering attempts.

Transformer-based natural language models, particularly bidirectional encoder representations from transformers and its derivatives, have simultaneously become the dominant tool for phishing detection precisely because of their capacity

Corresponding Author: Paul Anderson, Maricopa County Community College District, Email: adn2120472@maricopa.ed

How to cite this article: Anderson, P. (2025). Beyond the Human Signature: Transformer-Based Detection of AI-Generated Phishing Emails and Social Engineering Attacks. *Journal of Science, Technology and Social Transformation* 1(4), 46-53.

Source of support: Nil

Conflict of interest: None

to model context and semantics rather than relying on superficial lexical cues. This creates an unusual situation in which the same underlying architecture, the transformer, sits on both sides of an emerging arms race: generating convincing deceptive text on one side and classifying it on the other. Understanding how well transformer-based detectors actually perform against transformer-generated attacks, as opposed to the legacy human-authored phishing corpora on which most published accuracy figures are based, is therefore a question of considerable practical urgency and, as this review will show, one that the literature has only begun to address systematically.

This paper pursues three research questions. First, what architectural and methodological patterns characterize the application of transformer-based models to phishing and social engineering detection. Second, to what extent do reported performance gains generalize from conventional, human-authored phishing corpora to messages produced or refined by large language models. Third, what complementary signals or design principles might close the gap between benchmark performance and real-world robustness against generative adversaries. The analysis draws conceptually on the elaboration likelihood model from social psychology, which explains why persuasive messages succeed or fail depending on the cognitive route through which a recipient processes them, and on the broader cybersecurity principle of defense in depth, which holds that no single detection layer, however capable, should be trusted as a sole safeguard. Together these frameworks motivate treating transformer-based classification as one component within a layered detection architecture rather than as a complete solution in itself.

The paper proceeds in six sections. Section two synthesizes the relevant literature across four thematic clusters: the evolution of phishing detection toward transformer architectures, comparative performance of encoder-based and distilled models, the emergence of large language models as both attack tool and detection target, and the persistent challenge of explainability and adversarial robustness. Section three describes the theoretical, integrative research design employed in this synthesis. Section four presents findings organized around a comparative mapping of model families, illustrated with a process diagram of a representative detection pipeline. Section five discusses the implications of these findings, and section six concludes with directions for future research.

Literature Review

From Feature Engineering to Contextual Transformers

Phishing email detection research began largely with handcrafted feature sets built around sender metadata, embedded links, and lexical markers, processed through classical classifiers such as naive Bayes, logistic regression, and support vector machines. These approaches remain computationally efficient and continue to perform respectably on curated benchmark datasets, but they depend on manually specified features that generalize poorly once attackers adjust their tactics. The introduction of deep learning architectures, and subsequently of transformer-based encoders, shifted the field toward models capable of learning contextual representations directly from raw text. A comprehensive evaluation spanning fourteen machine learning and deep learning models across ten datasets found that transformer-based architectures, specifically BERT and RoBERTa, achieved the highest detection accuracies of the entire comparison, exceeding ninety-eight and ninety-nine percent respectively on a merged and balanced corpus, outperforming traditional machine learning approaches by a meaningful margin. Critically, however, the same study found that when these top-performing models were subsequently tested on an external, previously unseen phishing dataset, their accuracy fell substantially below the figures obtained on the training-adjacent benchmark, a gap the authors attributed partly to the limitations of the next sentence prediction mechanism embedded in BERT's original pretraining objective.

Complementary work applying BERT specifically to the classification of phishing, spam, and legitimate email demonstrated strong performance across balanced and imbalanced dataset conditions, while emphasizing that the interpretability of the resulting model remained limited without dedicated explainable artificial intelligence techniques layered on top. Similarly, a study applying the BERT transformer architecture within a conference-length evaluation confirmed that contextual embeddings meaningfully outperformed earlier bag-of-words representations for phishing text classification, a pattern echoed by subsequent work exploring transformer-based identification of the specific persuasion principles, such as authority, urgency, and scarcity, embedded within phishing message content. Another study applying BERT directly to the detection of social engineering text argued that the model's capacity to encode long-range contextual dependencies allowed it to recognize manipulative framing that classical bag-of-words classifiers systematically missed.

Distilled and Efficient Transformer Variants

A parallel research thread addresses the practical cost of deploying full-scale transformer models within real-time email security gateways. DistilBERT, introduced as a compressed variant retaining the majority of BERT's language modeling performance while reducing model size and inference latency substantially, has been repeatedly identified in the reviewed literature as the most favorable candidate for production deployment precisely because prediction speed matters as much as raw accuracy in an operational security pipeline. Comparative benchmarking work confirms that although BERT and RoBERTa consistently edge out DistilBERT on accuracy, DistilBERT's markedly lower prediction time

makes it the more practical choice where message volume is high and latency budgets are tight. This trade-off between marginal accuracy gains and computational cost recurs throughout the literature and suggests that the optimal transformer variant for phishing detection is not a fixed answer but a function of deployment context, an observation with direct relevance for the argument developed later in this paper regarding layered defense.

Large Language Models as Both Threat and Countermeasure

The most consequential recent development in this literature concerns the dual role that generative large language models now occupy. Empirical work has directly demonstrated that commercial conversational models can be prompted, sometimes with only modest adversarial effort, to produce convincing phishing emails and related social engineering content, a finding that has been independently corroborated by subsequent research examining the ease with which content moderation safeguards in such systems can be circumvented for this purpose. Complementary research from cybersecurity practitioners has extended this concern to a systematic empirical study of how large language models can be used to devise, and separately to detect, phishing content, concluding that while generative models measurably increase the persuasiveness and personalization of phishing attempts, they can simultaneously be repurposed as detection tools when prompted to evaluate suspicious text, albeit with calibration behavior that differs meaningfully from purpose-built classifiers.

This dual-use dynamic has motivated a distinct body of scholarship on detecting large language model generated text as a general problem, independent of any specific application to phishing. Surveys of this literature describe a shift from early machine-generated text detectors, which relied on relatively simple statistical or stylometric signals, toward more sophisticated approaches addressing adversarial robustness and cross-domain generalization following the widespread public availability of highly capable conversational models. Stylometric and perplexity-based detection methods, which quantify the statistical predictability or burstiness of a passage of text under a reference language model, have shown genuine forensic value, yet the same body of work cautions that these signals degrade considerably once generated text is edited, paraphrased, or blended with human-authored content, a scenario increasingly typical of real-world social engineering campaigns in which attackers use generative tools as a drafting aid rather than a fully automated pipeline. A specifically security-focused system built to leverage large language models for phishing email detection reinforces this picture, reporting strong performance on held-out phishing samples while explicitly noting that detection accuracy declines when the input text departs from the stylistic patterns represented in training data, underscoring that the underlying generalization problem observed in general text classification research reappears in the security-specific application.

Explainability, Robustness, and the Limits of a Purely Linguistic Signal

As with other applications of deep learning to security, the opacity of transformer-based phishing classifiers has drawn sustained scholarly attention. Explainable approaches applying post hoc techniques to fine-tuned transformer models have been proposed specifically to surface which words or sub-word tokens most influenced a given phishing classification, a capability treated in the literature as necessary rather than optional given the operational requirement that human analysts be able to audit and trust automated flags before acting on them. Separately, empirical work analyzing the robustness of BERT-based spam and phishing classifiers under adversarial perturbation has shown that carefully crafted text modifications, including synonym substitution and minor structural alteration, can meaningfully degrade classifier performance, a vulnerability of particular concern given that generative language models can automate precisely this kind of adversarial rewriting at scale.

Taken together, this cluster of literature converges on a conclusion with significant implications for the field: linguistic fluency and grammatical correctness, the very qualities that generative language models supply so readily, were never in themselves reliable markers of legitimacy, and their disappearance from phishing content does not eliminate detectability so much as it shifts the burden of detection toward behavioral, contextual, and structural signals that operate independently of surface language quality, including sender reputation, embedded link characteristics, timing anomalies, and organizational context. Recent comparative studies of phishing detection approaches that incorporate metadata and uniform resource locator features alongside textual content report meaningfully improved robustness relative to text-only classifiers, lending empirical weight to this argument. The gap this paper addresses is the absence, within the reviewed literature, of a synthesized account connecting the technical performance of transformer-based text classifiers to this broader argument about the diminishing sufficiency of linguistic signal alone, a connection that becomes especially urgent as large language models continue to erode the traditional linguistic markers of deception.

Methodology

This study employs a qualitative, theoretical research design built around an integrative literature synthesis rather than an original empirical experiment. The choice of design follows directly from the paper's aim, which is to consolidate and critically interpret a rapidly expanding and thematically fragmented body of scholarship on transformer-based phishing detection and large language model generated

text, in order to surface patterns and contradictions that no single primary study, constrained to its own dataset and architecture, is positioned to reveal. This approach is well established in information systems and cybersecurity scholarship, particularly for subfields experiencing the kind of rapid publication growth documented throughout the reviewed literature.

The literature corpus was assembled through structured searching of peer-reviewed journals, established conference proceedings, and widely cited preprint repositories, using search terms combining transformer, BERT, RoBERTa, large language model, phishing, social engineering, and AI-generated text detection. Sources were selected according to three criteria: peer-reviewed or otherwise well-established publication venue, direct relevance to transformer-based or large language model based detection of deceptive text, and recency, with particular emphasis on work published from 2022 onward to capture the post-ChatGPT shift in the threat landscape while retaining foundational architectural papers necessary for technical grounding. This narrative-integrative approach, rather than a fully systematic review with rigid inclusion algorithms, was chosen because the central research questions require interpretive judgment about thematic relevance across a literature that spans natural language processing, cybersecurity, and social psychology, a breadth that predefined systematic protocols are not well suited to accommodate.

Analytical procedure followed an iterative thematic coding process. Each source was examined for its stated research problem, model architecture, dataset composition, reported performance metrics, and any limitations the original authors themselves acknowledged, particularly regarding generalization beyond the training distribution. These elements were organized into the four thematic clusters presented in the literature review, and recurring cross-cutting patterns, most notably the tension between high benchmark accuracy and degraded out-of-distribution performance, were identified through constant comparison across the assembled sources. The conceptual comparison of model families presented in the findings section was developed inductively from this coding process, constructed by identifying which combinations of architecture, strength, and reported weakness recurred most consistently across the literature rather than being derived from any single source.

Several considerations bear on the validity of this synthesis. Because the paper synthesizes secondary literature rather than analyzing primary data, its principal validity concern is the representativeness of the assembled corpus. This was addressed by deliberately including sources spanning divergent methodological traditions, from applied benchmarking studies reporting quantitative accuracy figures to conceptual and survey papers addressing detection theory more broadly, and by cross-referencing claims across independent sources before treating a pattern as representative rather than idiosyncratic. Reliability in the qualitative sense was supported by returning to primary sources whenever thematic classification proved ambiguous rather than relying on a single reading. Ethically, the paper takes particular care not to overstate the operational readiness of any detection technique discussed, a concern of direct practical significance in a domain where premature confidence in an automated safeguard can itself become a vulnerability, and it deliberately omits any specific prompt engineering detail that could meaningfully lower the barrier to generating convincing phishing content, consistent with the legitimate scope of an academic synthesis focused on defensive capability rather than offensive technique.

Results and Findings

The synthesis yields a comparative mapping of five model families used in transformer-based phishing and social

Table 1: Comparative Mapping of Model Families for Transformer-Based Phishing and Social Engineering Detection

Model family	*Representative architecture*	*Reported strength*	*Reported weakness*
Encoder-only transformers	BERT, RoBERTa, DistilBERT	High accuracy on benchmark corpora; strong contextual disambiguation	Costly fine-tuning; degraded accuracy on unseen phishing campaigns
Distilled/compressed transformers	DistilBERT, TinyBERT variants	Lower latency; feasible for real-time email gateways	Modest accuracy trade-off relative to full-size models
Hybrid neural architectures	BERT-LSTM, CNN-transformer ensembles	Captures both sequential and contextual cues	Increased architectural complexity and interpretability burden
Large language model detectors	GPT-family classifiers, zero-shot prompting	Adaptable to novel, LLM-crafted phishing without retraining	Inconsistent calibration; vulnerable to adversarial paraphrase
Stylometric/perplexity detectors	Log-likelihood and burstiness scoring	Model-agnostic; does not require labeled phishing corpus	Weakens sharply once text is human-edited or paraphrased

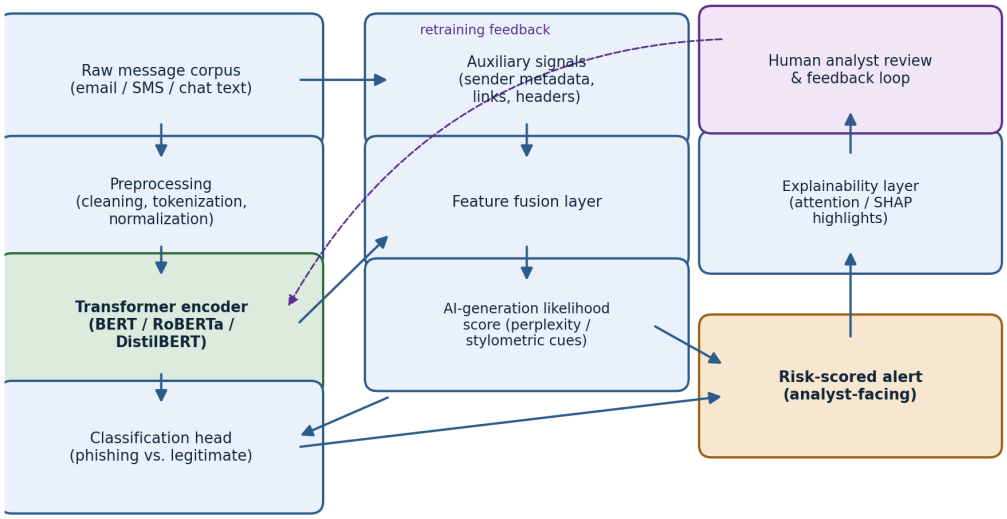


Figure 1: Representative transformer-based detection pipeline with auxiliary signal fusion, explainability layer, and human-in-the-loop feedback.

engineering detection, summarizing the architecture each family represents, its principal reported strength, and its principal reported weakness as documented across the reviewed literature. This mapping is presented in Table 1.

Three patterns stand out from this comparison. First, encoder-only transformers and their distilled variants dominate reported accuracy figures on conventional benchmark corpora, but the margin separating them from earlier deep learning architectures such as long short-term memory networks is often narrower than the attention such models receive would suggest, and the accuracy advantage narrows further, sometimes substantially, once evaluation moves to out-of-distribution or externally sourced test data. Second, large language model based detectors occupy a distinctive niche precisely because they can be deployed against novel phishing campaigns without task-specific retraining, a genuine advantage given how quickly generative attackers can iterate on message style, but this flexibility comes paired with calibration behavior that is less predictable and less thoroughly benchmarked than purpose-built classifiers. Third, stylometric and perplexity-based approaches, though conceptually elegant and appealing precisely because they require no labeled phishing corpus, are the family most acutely vulnerable to the specific threat this paper is concerned with, since even modest human editing of machine-generated text measurably degrades the statistical signal on which they depend.

Figure 1 illustrates a representative detection pipeline synthesized from the architectures described across the reviewed literature. Raw message text is preprocessed and passed through a transformer encoder to obtain contextual embeddings, which are combined with auxiliary metadata signals through a fusion layer; a parallel channel estimates the likelihood that the message content was machine-generated using perplexity or stylometric cues, and both streams inform a risk-scored alert that is surfaced to a human analyst together with an explainability layer highlighting the tokens or features that drove the classification. The dashed feedback path denotes the retraining loop by which analyst adjudications are fed back into the model, a mechanism the reviewed literature consistently identifies as necessary for maintaining detection accuracy against an adversary that adapts its tactics over time.

A further finding concerns the empirical evidence, still comparatively sparse but growing, on detection performance against large language model generated phishing specifically rather than legacy human-authored corpora. Where such evaluations exist, they consistently report a measurable accuracy decline relative to performance on conventional benchmarks, and several studies attribute this decline specifically to the absence of the surface-level anomalies that earlier classifiers, whether explicitly or implicitly, learned to rely upon during training. This finding is significant because it indicates that reported benchmark accuracy, however impressive in isolation, should not be read as a reliable indicator of resilience against the specific and rapidly growing threat that motivated this synthesis in the first place.

Discussion

These findings carry direct implications for both research design and operational practice. For research, the consistent gap between benchmark accuracy and out-of-distribution performance suggests that future studies in this area should treat evaluation against large language model generated content, rather than only against legacy human-authored corpora, as a standard rather than an optional robustness

check, since the current practice of reporting benchmark figures without such testing risks materially overstating readiness for deployment. For practice, the finding that stylometric and perplexity signals degrade under light human editing implies that organizations should not rely on any single detection modality, however theoretically appealing, and should instead favor layered architectures of the kind depicted in Figure 1, in which linguistic, behavioral, and contextual signals are fused and routed through human review rather than acted upon automatically.

Situated within the theoretical frameworks introduced earlier, these findings extend the elaboration likelihood model into the technical domain by suggesting that generative language models primarily strengthen the peripheral persuasive cues, fluency, tone, and apparent legitimacy, that drive low-effort message processing, which is precisely the route through which most successful phishing operates; detection systems that focus solely on linguistic anomaly are therefore attacking exactly the surface that generative models have made least reliable. The defense-in-depth principle, meanwhile, receives direct empirical support from the finding that no single model family in Table 1 dominates across both accuracy and robustness, reinforcing the case for combining complementary detection modalities rather than searching for a single superior architecture.

The principal limitation of this synthesis follows from its method: reliance on secondary literature means the paper cannot adjudicate between studies using incompatible datasets or evaluation protocols, and it may underrepresent proprietary detection systems developed within industry that are not published in academic venues. Future research would benefit from standardized adversarial benchmarks that explicitly include large language model generated and human-edited hybrid phishing samples, from longitudinal studies tracking detector performance as generative models themselves continue to evolve, and from more rigorous study of human-in-the-loop configurations to determine which explainability presentations most effectively support analyst judgment without inducing either excessive automation trust or alert fatigue. As regulatory attention to high-risk artificial intelligence systems continues to expand globally, research in this area will also need to address questions of accountability when automated phishing detection systems fail, alongside its traditional focus on raw classification performance.

Conclusion

This paper has synthesized the literature on transformer-based detection of phishing emails and social engineering attacks, with particular attention to the emerging challenge posed by large language model generated content, and has developed a comparative framework mapping the strengths and failure modes of the principal model families used in this domain. The synthesis demonstrates that transformer architectures deliver genuine and well-documented gains over earlier detection approaches, yet these gains are measured predominantly against benchmark corpora that no longer fully represent the threat landscape they are meant to model, and performance degrades in ways the field has only begun to quantify once generative adversaries enter the picture. The central scholarly contribution of this paper lies in connecting these previously separate strands, transformer benchmarking research and large language model text detection research, into a single account of why linguistic fluency can no longer serve as the primary signal of deception. Future work stands to benefit substantially from adversarial benchmarking that treats generative attacks as the norm rather than the exception, and from detection architectures that, in keeping with the defense-in-depth principle, extend deliberately beyond text alone.

References

[1] Al-Subaiey, A. A., Al-Thani, A. M., Alam, N. A., Fatema Antora, K., Khandakar, A., & Uz Zaman, S. M. A. (2024). Novel interpretable and robust web-based AI platform for phishing email detection. Computers & Electrical Engineering, 120, 109625. https://doi.org/10.1016/j.compeleceng.2024.109625

[2] Alhuzali, A., Alloqmani, A., Aljabri, M., & Alharbi, F. (2025). In-depth analysis of phishing email detection: Evaluating the performance of machine learning and deep learning models across multiple datasets. Applied Sciences, 15(6), 3396. https://doi.org/10.3390/app15063396

[3] Altwaijry, N., Al-Turaiki, I., Alotaibi, R., & Alakeel, F. (2024). Advancing phishing email detection: A comparative study of deep learning models. Sensors, 24(7), 2077. https://doi.org/10.3390/s24072077

[4] Chen, Z., Zhao, L., & Yin, G. (2026). A predictive deep learning model for cyber threat detection in a smart city environment. Security and Privacy, 9(3), e70218.

[5] Routhu, K. K. (2022). From RFID to Geofencing: IoT-Enabled Smart Time Tracking in Oracle HCM Cloud. *International Journal of Science, Engineering and Technology*, *10*(4).

[6] Suresh, A. (2024). Auto-BI Frameworks Powered by Generative Artificial Intelligence for Scalable Self-Service Data Analytics in Large Organizations. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 5(3), 191-201. https://doi.org/10.63282/3050-9262.IJAIDSML-V5I3P119

[7] Padur, S. K. R. (2024). AI-augmented platform engineering: Redefining developer experience through autonomous, self-optimizing enterprise systems. *International Journal of Science, Engineering and Technology*.

[8] Routhu, K. K. (2024). Beyond Automation: AI-Powered Employee Engagement Journeys in Oracle HCM Cloud. *KOS Journal of AIML, Data Science, and Robotics*, *1*(1), 1-6.

[9] Padur, S. K. R. (2024). Securing Oracle Integration Cloud ERP ecosystems, zero trust architecture, data governance, and compliance automation. *International Journal of Science, Engineering and Technology*, *12*(4), 10-5281.

[10] Routhu, K. K. (2024). The future of HCM: Evaluating Oracle's and SAP's AI-powered solutions for workforce strategy. *Journal of Artificial Intelligence, Machine Learning & Data Science*, *2*(2), 2942-2947.

[11] Padur, S. K. R. (2024). Designing Scalable AI Architectures: Foundations for Enterprise-Grade Intelligence Systems. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol*, *10*(7), 274-278.

[12] Padur, S. K. R. (2023). AI-Augmented Enterprise ERP Modernization: Zero-Downtime Strategies for Oracle E-Business Suite R12. 2 and Beyond. *Available at SSRN 5605510*.

[13] Routhu, K. K. (2023). AI-driven succession planning in Oracle HCM Cloud: Building resilient leadership pipelines through predictive analytics. *International Journal of Science, Engineering and Technology*, *11*(5).

[14] Padur, S. K. R. (2022). AI augmented platform engineering, transforming developer experience through intelligent automation and self optimizing internal platforms. *International Journal of Science,*

Engineering and Technology, *10*(5), 10-5281.
[15] Padur, S. K. R. (2022). Intelligent resource management: AI methods for predictive workload forecasting in cloud data centers. *J. Artif. Intell. Mach. Learn. & Data Sci*, *1*(1), 2936-2941.
[16] Routhu, K. K. (2023). Embedding fairness into the digital enterprise: Data-driven dei strategies with oracle hcm analytics. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol*, *9*(8), 266-274.
[17] Routhu, K. K. (2022). From Case Management to Conversational HR: Redefining Help Desks with Oracle's AI and NLP Framework. *International Journal of Science, Engineering and Technology*, *10*(6).
[18] Padur, S. K. R. (2021). Deep learning and process mining for ERP anomaly detection: Toward predictive and self-monitoring enterprise platforms. *Available at SSRN 5605531*.
[19] Routhu, K. K. (2020). Strategic compensation equity and rewards optimization: A multi-cloud analytics blueprint with Oracle Analytics Cloud. *Available at SSRN 5737266*.
[20] Padur, S. K. R. (2017). Engineering resilient datacenter migrations: Automation, governance, and hybrid cloud strategies. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN*, 2456-3307.
[21] Routhu, K. K. (2021). Harnessing AI Dashboards in Oracle Cloud HCM: Advancing Predictive Workforce Intelligence and Managerial Agility. *International Journal of Scientific Research & Engineering Trends*, *7*(6).
[22] Padur, S. K. R. (2016). Network modernization in large enterprises: Firewall transformation, subnet re-architecture, and cross-platform virtualization. *International Journal of Scientific Research & Engineering Trends*, *2*(5).
[23] Shyam, D., Manoharan, P., Syed, M. H., Ragireddy, U. K., Addepalli, P. V., & Bandaru, S. R. (2024). Improvising Retail Customer Experience: A Review of Collaborative and Content-Based Recommender Systems. *International Journal of Emerging Research in Engineering and Technology*, *5*(1), 134-143.
[24] Addepalli, P. V., Bandaru, S. R., Shyam, D., Manoharan, P., Syed, M. H., & Ragireddy, U. K. (2024). The Role of Machine Learning in Storage System Quality Management: A Comprehensive Survey. *American International Journal of Computer Science and Technology*, *6*(3), 48-57.
[25] Bandaru, S. R., Shyam, D., Manoharan, P., Syed, M. H., Ragireddy, U. K., & Addepalli, P. V. (2024). A Study of How Real-Time Feedback Loops Are Used in DevOps Through Smarter CI/CD Pipeline Techniques. *International Journal of Emerging Trends in Computer Science and Information Technology*, *5*(2), 164-173.
[26] Syed, M. H., Ragireddy, U. K., Addepalli, P. V., Bandaru, S. R., Shyam, D., & Manoharan, P. (2024). Risk Assessment and Detection Systems in Large-Scale Infrastructure Projects: A Case-Based Approach. *International Journal of AI, BigData, Computational and Management Studies*, *5*(2), 192-201.
[27] Manoharan, P., Syed, M. H., Ragireddy, U. K., Addepalli, P. V., Bandaru, S. R., & Shyam, D. (2024). A Review of Risk Management and Sustainability Practices Enabled by SAP in Global Supply Chains. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *5*(4), 266-275.
[28] Mylavarapu, K. B., Pilli, J., Jannu, P. K., Mohammad, J. A., & Panchali, S. H. (2024). Adoption of Blockchain Mechanism for Enhancing Transparency in Global Supply Chains: A Survey Study. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *5*(1), 192-199.
[29] Mohammad, J. A., Panchali, S. H., Mylavarapu, K. B., Pilli, J., & Jannu, P. K. (2024). Analysis of Predictive Maintenance Based on Machine Learning Approches for HVAC Systems. *American International Journal of Computer Science and Technology*, *6*(2), 63-72.
[30] Panchali, S. H., Mylavarapu, K. B., Pilli, J., Jannu, P. K., & Mohammad, J. A. (2024). Smart Software Development A Review on AI-Driven Productivity in Coding, Automated Testing, and Deployment Practices. *International Journal of Emerging Trends in Computer Science and Information Technology*, *5*(1), 172-181.
[31] Mylavarapu, K. B., Pilli, J., Jannu, P. K., Mohammad, J. A., & Panchali, S. H. (2024). Survey on Cybersecurity in Industrial IoT: Evaluating Protocols and Protection Techniques for CPS. *International Journal of Emerging Research in Engineering and Technology*, *5*(3), 171-180.
[32] Pilli, J., Jannu, P. K., Mohammad, J. A., Panchali, S. H., & Mylavarapu, K. B. (2024). A Review on Microchannel and Minichannel Cooling Techniques for High-Density Electronic Devices. *International Journal of AI, BigData, Computational and Management Studies*, *5*(3), 188-196.
[33] Chilakapati, V. K., Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., & Dasari, A. (2024). Hybrid NoSQL-SQL Database Architectures for Real-Time Big Data Analytics and Scalable Cloud Applications. *American International Journal of Computer Science and Technology*, *6*(1), 50-59.
[34] Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., Dasari, A., Chilakapati, V. K., & Keshireddy, S. R. (2024). Early Diseases Diagnosis of Chropnic via Machine Learning Based Models in Big Data Health Records. *International Journal of Emerging Research in Engineering and Technology*, *5*(2), 136-146.
[35] Kavuluri, H. V. R., Pathani, A. K., Dasari, A., Chilakapati, V. K., Keshireddy, S. R., & Nagumotu, V. T. (2024). Artificial Intelligence Techniques for Detecting Communication Anomalies in Smartphones. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *5*(3), 202-210.
[36] Taluri, R. (2024). A Cloud-Native Reference Architecture for Data Engineering, Generative AI, and Decision Intelligence Using AWS and Amazon Bedrock. International Journal of AI, BigData, Computational and Management Studies, 5(1), 218-227. https://doi.org/10.63282/3050-9416.IJAIBDCMS-V5I1P122
[37] Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., Dasari, A., & Chilakapati, V. K. (2024). Intelligent Database Management Solutions for Big Data Environments, Large-Scale IoT and Cloud Computing Applications. *International Journal of Emerging Trends in Computer Science and Information Technology*, *5*(4), 156-165.
[38] Pathani, A. K., Dasari, A., Chilakapati, V. K., Keshireddy, S. R., Nagumotu, V. T., & Kavuluri, H. V. R. (2024). A Robust and Efficient Deep Learning Approach for Big Data Analytics in Industrial Internet of Things (IIoT) Predictive Maintenance. *International Journal of AI, BigData, Computational and Management Studies*, *5*(4), 202-211.
[39] Suresh, A. (2024). Auto-BI Frameworks Powered by Generative Artificial Intelligence for Scalable Self-Service Data Analytics in Large Organizations. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 5(3), 191-201. https://doi.org/10.63282/3050-9262.IJAIDSML-V5I3P119
[40] Taluri, R. (2024). A Hybrid Data Engineering and Generative AI Architecture for Intelligent Data Governance, Metadata Management, and Automated Data Quality Assessment on AWS. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 5(2), 230-240. https://doi.org/10.63282/3050-9262.IJAIDSML-V5I2P126
[41] Bandaru, S. R., Shyam, D., Manoharan, P., Syed, M. H., Ragireddy, U. K., & Addepalli, P. V. (2023). Data-Driven Approaches for Forecasting Cost Overruns in Infrastructure Projects. *European Journal of Technology*, *7*(5), 23-34.
[42] Dhuli, S., Prabu, M., Muzaffer, H. S., Uday Kumar, R., & Prasanth Varma, A. (2023). A Survey on Critical Success Factors Influencing Project Success Across Multiple Industries. *J Contemp Edu Theo Artific Intel: JCETAI-106*.
[43] Ragireddy, U. K., Addepalli, P. V., Bandaru, S. R., Shyam, D., Manoharan, P., & Syed, M. H. (2023). An Overview of Different Automation Strategies Used in Cloud-Based CI/CD Pipelines for Software Deployment. *International Journal of Technology, Management and Humanities*, *9*(01), 116-127.
[44] Addepalli, P. V., Bandaru, S. R., Shyam, D., Manoharan, P., Syed, M. H., & Ragireddy, U. K. (2023). Performance Analysis of Machine Learning Models for Cybersecurity Risk Classification in IT Projects. *Universal Library of Engineering Technology*, (Issue).
[45] Mylavarapu, K. B., Pilli, J., Jannu, P. K., Mohammad, J. A., & Panchali, S. H. (2023). Artificial Intelligence for Continuous Quality Assurance in Cloud-Native Applications: A Review. *European Journal of Technology*, *7*(5), 35-49.
[46] Jannu, P. K., Mohammad, J. A., Panchali, S. H., Mylavarapu, K. B., & Pilli, J. (2023). A Survey on Energy-Efficient Microarchitecture Design Strategies for Mobile CPUs and GPUs: Techniques and Future Directions. *International Journal of Technology, Management and Humanities*, *9*(04), 137-151.
[47] Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., Dasari, A., & Chilakapati, V. K. (2023). An AI-Based Framework for Detecting

IoT Botnets Through Network Traffic Analysis and Modeling. *Universal Library of Engineering Technology*, (Issue).

[48] Kavuluri, H. V. R., Pathani, A. K., Dasari, A., Chilakapati, V. K., Keshireddy, S. R., & Nagumotu, V. T. (2023). Cloud Data-Center Network Verification: Approaches, Algorithms and Toolchains. *American Journal of Computing and Engineering*, *6*(2), 18-32.

[49] Chilakapati, V. K., Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., & Dasari, A. (2023). A Survey on 5G/6G Radio Access Networks: Technologies, Protocols, and Performance Metrics. *International Journal of Technology, Management and Humanities*, *9*(02), 79-93.

[50] Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., Dasari, A., & Chilakapati, V. K. (2023). An AI-Based Framework for Detecting IoT Botnets Through Network Traffic Analysis and Modeling. *Universal Library of Engineering Technology*, (Issue).

[51] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT 2019, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[52] Otieno, D. D., Namin, A. S., & Jones, K. S. (2023). The application of the BERT transformer model for phishing email classification. In Proceedings of the IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC) (pp. 1303–1310). IEEE. https://doi.org/10.1109/COMPSAC57700.2023.00198

[53] Pai, Y., & Suresha, D. (2026). A comprehensive predictive information security risk intelligence framework for IT service environments: An empirical machine learning approach. International Journal of Advances in Signal and Image Sciences, 1022–1041.