Springer

SEET 2026

PennState Behrend

ABSTRACT BOOK

2ND INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING OF EMERGING TECHNOLOGIES

August 10 - 11, 2026
Erie, Pennsylvania, USA

Springer

SEET 2026

PennState Behrend

ABSTRACT BOOK

2ND INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING OF EMERGING TECHNOLOGIES

August 10 - 11, 2026
Erie, Pennsylvania, USA

CONTENT

I.D#	Paper Title	Author Name	P#
71	Empirical Investigation of Cyber Attack Detection	Joshua Partridge, Alex Wood, Ghulam Mudassir	55
84	Biologically Plausible Learning in Spiking Neural Networks Using Indirect Feedback Alignment: A study of the implementation of Feedback Alignment methods into Spiking Neural Networks	Joshua Nallaraja and Jonathan Liaw	56
95	Enhancing Requirement Analysis with Hybrid AI-Based Question Answering Systems	Yongyan Liang⋆, Lihua Lan, Muhammad Abdul Basit Ur Rahim	57
	An Exploratory Machine Learning Approach for Mental-Health Related Text Classification	Brandon Breindel and Muhammad Talha Gul	58
117	A Locally Deployed AI Assistant with Natural Language and Symbolic Computation Capabilities	Zaheen Chowdhury and Muhammad Abdul Basit Ur Rahim	59
	Analytics Framework for Course Engagement and Evaluation	Luz Diaz O'Keefe, Raquel Anden Rhoads, Dusan Ramljak	60
74	Hyperdimensional Computing on Apple Neural Engine Chips	Srivatsan Suresh Babu, Sophia Shahryar, and Justin Morris	61
Ph.D RESEARCH			
88	Lightweight Data Lineage Tracing for Cross-Function Leak Detection in Serverless Applications	Zuhaib Nishter and Gang Li	64
148	Epic-Organized vs. Requirement-Aligned Gherkin: An Empirical Evaluation of LLM-Based Acceptance Criteria Generation	Shahbaz Siddeeq , Mateen Abbasi , Jussi Rasku , Zheying Zhang , François Christophe , Tommi Mikkonen , and Pekka Abrahamsson	65
164	An Overview of Human Aspects in the Alignment of Requirements Engineering and Testing	Afsarah Jahin, Iflaah Salman, Rahul Mohanani, and Burak Turhan	66
165	Dynamic Trustworthiness of AI for Autonomous Driving: A Design Problem Definition	Daupadie Laknamali Ranatunga, Vishaka Basnayake, Nada Elgendy, and Tero Paivarinta	67
167	Governing LLM-Based Adaptive Components in Safety-Critical Clinical Decision Support: A Constraint-Layer Extension of MAPE-K	Kavishwa Wendakoon1 and Nirnaya Tripathi	68
168	HyQ-Guard – AI-Driven Runtime Anomaly Monitoring and Mitigation for Hybrid Quantum Workflows	Ejaz Ahmed, Mikko Mäkitalo	69
169	Operationalising the EU AI Act in Software Engineering: A Systematic Mapping Study with a Focus on Healthcare	Qaiser Khan, Abdul Raffay Saeed, and Rahul Mohanani	70
171	Triangulating Prompt Engineering Effectiveness for Edge Case Detection: Integrating Expert Quality Review with Association Rule Mining in LLM-Driven Software Testing	Usama Azhar, Saif Ur Rehman Khan, Shahid Hussain	71

I.D#	Paper Title	Author Name	P#
177	Reliability-Aware Retinal Encoder for Ordinal Diabetic Retinopathy Severity Modelling	Sameera Gamage	72
178	A Reproducible Multi-View Mammogram Classification Baseline Using Dual-View Fusion on VinDr-Mammo	Jitendra Sai Kota, Saddam Hossain Irfan, and Muhammad Imran	73
192	Supporting Reflection on Sustainable Software Design Decisions in Software Engineering Education: An Evaluation of a Socratic Chatbot	Kseniia Suomalainen, Elena Cavallin, Viktoriia Olshanskaia1, Maria Luce Lupetti, and Jari Porras	74
130	Analytics Framework for Course Engagement and Evaluation	Luz Diaz O'Keefe, Raquel Anden Rhoads, Dusan Ramljak	75
132	Autonomous Multi-Agent Governance: AI Sentinel Framework for Mitigating Psychological Restraints in AI-Driven Fall Management	Tushar, Anushka Thagle1, Supanut Chindawan, Sanjna Sai Chintakunta, Muhammad Faizan Raza, Albert Elcock, Rich Moore, Praneeth Sunkavalli, Chandan Shivalingaiah, Gaurav Prabhu, Lukas Sparer, Alexander Neulinger, Sanjeev Sondur, Arturo Casasa, Maryam Roshanaei, and Dušan Ramljak	76

Regular Research

Dynamics of Artificial Intelligence and Its Impact on Modern Higher Education

Author Name: Rashid Ali Khan (PhD)

Abstract: This study aims to explore why university students begin using AI tools. It applies Rogers' Diffusion of Innovation theory as a framework to examine the key factors influencing students' decisions to adopt AI for learning purposes. Specifically, the study focuses on four main factors: how useful students perceive the tools to be, how easy they find them to use, the influence of those around them, and how seamlessly the tools fit into their existing study habits. A quantitative methods approach is used—combining surveys with interviews—to gather both quantitative data and personal insights. The results are anticipated to provide valuable guidance for educators, policymakers, and technology developers seeking to promote the responsible integration of AI in higher education. In the end, the study seeks to deepen understanding of how emerging technologies like AI become embedded in academic settings.

Driving Governance and Adoption in Custom Faculty Workload System Integrations within Cloud ERP for Community Colleges

Sanjiv Kumar Bhagat, PMP, CSM, PROSCI

Abstract: Community colleges with shared-services PeopleSoft use cases have historically seen multiple iterations of missed part-time teaching workload functionality because ERP architectures designed for full-time faculty do not always accommodate variable load units, term-based contracts, and/or complex compensation schedules. The FWL bolt-on creates an extension layer built upon PeopleTools that retains all delivered core PeopleSoft functionality, adds a parallel data model, multi-step workflow approvals through PeopleSoft's Approval Workflow Engine, and tightly governed integration touchpoints with Human Capital Management, Campus Solutions, and Payroll for North America. For institutional data, the security architecture is layered, role-based, and scoped to multi-college consortia boundaries. Data governance, in three tiers (standards and procedures plus configurable and local exceptions), parallels the federal theory of information technology governance and the COBIT framework for the governance of enterprise IT. This distinguishes governance from management. The change management layer consists of the PROSCI/ADKAR model, which maps the elements of individual change to the primary interventions required to deploy awareness campaigns, champion networks, role-based training, hands-on practice environments, and post-deployment sustainment. These three layers of architecture, governance, and change management combine to provide a blueprint for the extension of ERP stacks in complex multinational, multi-institutional higher education contexts where technical realization alone is not sufficient to ensure operational success.

A Gate-Based Quantum Ising Model Approach to Solving KenKen Puzzles Using Binary Encoding

Author Name: Wen-Li Wang, Mei-Huei Tang, Kevin Wang, Shahid Hussain, Saif Ur Rehman Khan

Abstract: The KenKen puzzle is a classic example of a Constraint Satisfaction Problem (CSP) and has been proven to be NP-complete. CSPs are inherently tied to combinatorial structures and constraints, and quantum algorithms that leverage the principle of superposition are particularly well suited to addressing such challenging problems. As an extension and more advanced variant of Sudoku, KenKen shares certain structural constraints; however, it also incorporates arithmetic operations that introduce additional complexity and require specialized insight. Consequently, novel methods are needed to address these mathematical characteristics, while techniques developed for other hard problems can still be applied to handle the shared constraints. Building on prior work, this study extends a gate-based quantum Ising model, originally developed to solve Sudoku by minimizing the Hamiltonian energy, to address the KenKen puzzle. To reduce the number of required logical qubits, the proposed approach continues to adopt a base-2 binary encoding rather than a one-hot representation. The development process begins with identifying the rules shared across puzzle types and formulating mathematical expressions for the additional KenKen-specific constraints. Observable operators for the quantum models are then constructed using Pauli operators, enabling the computation of expectation values associated with individual rules. These operators are subsequently integrated into a complete KenKen solver designed for global optimization. The solver's validity is evaluated using sample puzzles implemented within the IBM Qiskit quantum framework, and the outcomes for 4 × 4 and 3 × 3 puzzles are presented and analyzed to confirm the correctness of the proposed approach.

Data Mesh as a Migration Strategy: A Federated Approach to Cloud-Native Data Engineering

Author Name: Richa Solanki

Abstract: Enterprise data modernization demands architectural, organizational, and governance transformation — not merely a technical infrastructure up-grade. Monolithic data warehouses and central ETL pipelines are structurally ill-equipped to handle the data volume, velocity, and variety of modern work-loads, and customary cloud migration approaches entail unacceptable technical debt or organizational overhead. Data Mesh is a decentralized and domain-oriented approach that frames cloud adoption as a continuing organizational transformation powered by four principles: domain-oriented ownership, data-as-a-product, self-serve data infrastructure, and Federated Computational Gov-ernance (FCG). It distributes data engineering ownership across business do-mains while adhering to enterprise-level standards through platform-based guardrails. This presents an alternative model to the centralized data lake or da-ta warehouse. In contrast to centralized migration programs, domain teams are empowered to migrate at a pace commensurate with their domain's complexity and commercial value. Data products are stable contracts between data produc-ers and consumers that decouple interfaces from implementations and are, therefore, not impacted by any backend infrastructure change. Platform engi-neering abstracts cloud complexity into self-service, standardized tooling, low-ering the barrier for domain teams to adopt the cloud. Federated governance au-tomates regulatory compliance within the deployment pipeline, enabling scala-ble governance of distributed portfolios of domain products and services, and eliminating the overhead of manual compliance review. The patterns of stran-gler fig replacement, hybrid coexistence, event-driven modernization, and data product wrapping give architects risk-calibrated ways to deliver incremental value. The primary capabilities that drive Data Mesh success at enterprise scale are federated governance automation, product discipline, and platform capability sequencing

AI-Driven Autonomous Data Pipeline Orchestration in Cloud-Native Enterprise Platforms

Author Name: Sudhir Saxena

Abstract - Context: Enterprise data platforms are rapidly transitioning to cloud-native distributed architectures, introducing significant operational complexity in orchestrating large-scale data pipelines at scale. **Problem:** Traditional rule-based schedulers and static workflow engines are insufficient to manage dynamic workloads, fluctuating resource availability, cost-performance trade-offs, and failure recovery in distributed cloud environments. **Method:** This article proposes an AI-driven autonomous orchestration framework that employs a five-agent multi-agent reinforcement learning model with a formally defined state space, action space, and multi-objective reward function to dynamically optimize task scheduling, resource allocation, and retry strategies. The framework integrates an LSTM-based predictive failure mitigation system trained on historical observability metrics and implements a self-healing closed-loop control architecture for automated remediation. **Results:** Simulation-based benchmarking against Airflow 2.6, Argo Workflows 3.4, and Kubeflow Pipelines 1.8 demonstrates a reduction in mean time to recovery from 48–52 minutes to 2.8 minutes, improvement in resource utilization from 54–61% to 83.6%, and SLA compliance improvement from 81–85% to 96.4% (all $p < 0.001$, Wilcoxon signed-rank test, $n = 50$ trials). **Conclusion:** The proposed framework establishes a unified theoretical and architectural foundation for autonomous enterprise data pipeline orchestration, demonstrating that AI-native control planes can substantially advance reliability, efficiency, and cost-effectiveness beyond static orchestration strategies.

Engineering and Evaluating a Machine-Learning Monitoring Pipeline for Rare-Event Early Warning: A Reference Architecture and a Four-Layer Evaluation Methodology

Author Name: Rui Zong, Yutong Zhao, Sen He, and Muhammad Abdul Basit Ur Rahim

Abstract: Machine-learning components inside production software systems are typically evaluated with classification metrics such as AUC, Brier score, or F1. The decision the deployed component actually makes is operational: an alarm fires or it does not, and standard classification metrics do not directly measure that quality. We propose a reference architecture for rare-event ML monitoring components and a four-layer evaluation methodology that combines classification quality, operational alarm quality, calm-period diagnostics, and stability under drift, with threshold provenance treated as a first-class artifact. The architecture and the methodology are exercised on a controlled case study in equitymarket stress monitoring using daily SPX, NDX, and RUT indices and VIX-family data with three forward-drawdown stress definitions, an expanding-window evaluation, and a six-model registry. Six named failure modes that are invisible to single-metric evaluation emerge from the case study and are surfaced by the proposed methodology. Four are surfaced by the unmodified pipeline artifacts: three through cross-layer metric disagreement (operational–classification disagreement, silent classification degradation under regime shift, and a calibration–ranking trade-off) and one through the validation summary, feature schema, and per-fold record (a data-quality cascade that empties the matrix-feature test window).

Two further architectural validation gaps — validator-bypass leakage and silent gaps in the validator rule set — require component-ablation and failure-injection contrasts the headline grid does not produce. A crosscutting empirical observation scopes the layer-4 reading: in the case-study setting, distributional drift is positively associated with operational lift in the case-study diagnostics, which we interpret as a cautionary Layer-4 signal rather than evidence that drift improves model quality. The cross asset replication is honest rather than confirmatory: the SPX-specific lift advantage of the matrix-feature learners does not survive on NDX or RUT under the shared minimal feature schema, which materially scopes the contribution. We also present component-ablation and failure-injection

Unified DT–FL–DRL–Blockchain Framework for MEC Offloading

Author Name: Anthony Gladdney, Shashi Bhushan Jha

Abstract: Task offloading in Mobile Edge Computing (MEC) is challenging in practice than most theoretical frameworks suggest. Edge servers proceed at uneven loads, users move constantly, and wireless channels shift in ways that are hard to predict. Analytical optimization methods lean to break down under these conditions, and purely online reinforcement learning methods are timeconsuming to converge because they need large exploration in live systems. Previous research has tackled pieces of this problem with digital twin-assisted learning, federated DRL (Deep Reinforcement Learning), blockchain-secured FL (Federated Learning), but integrating all four mechanisms together in a single coherent design is missing in the existing literature. This research proposes a closed-loop architecture that integrates Digital Twin (DT) simulation, Multi-Agent Proximal Policy Optimization (MAPPO), FL, and blockchain verification for MEC task offloading. The DT allows agents to evaluate decisions in a virtual environment before to act on the real system. MAPPO manages decentralized policies with a shared critic. FL combines server experiences every three iterations through reputation weighted Federated Average (FedAvg), and a consortium blockchain filters low-quality updates while preserving a transparent trust record. In a simulated environment with 10 mobile users and 4 edge servers, the framework converges 43% faster than a non-federated PPO (Proximal Policy Optimization) baseline, gains a 14.7% higher peak reward, and carries blockchain reputation from 0.70 to a stable range of [0.920, 1.000] by iteration 50 with one random seed.

Applying Metamorphic Testing to Machine Learning: A Systematic Mapping Study

Author Name: A. Redempta Manzi Muneza, Kaveen Liyanage, Faqeer Ur Rehman, Emma Sheppard, and Ann Marie Reinhold

Abstract: A reliable test oracle. Metamorphic testing (MT) addresses this challenge by defining expected relationships between multiple inputs and their outputs, called metamorphic relations. Violations of these relations indicate potential faults. Existing publications demonstrate effectiveness of MT in evaluating ML models. This article presents a systematic mapping study of MT applied to ML over the past decade. We analyze publication trends, venues, countries, problem types, learning paradigms, algorithms, and application domains. The results show steady growth in MT research, with a strong focus on supervised learning, classification tasks, and neural network–based models, particularly in computer vision and natural language processing. However, only a few publications focused on cybersecurity and cyber-physical systems applications, indicating important areas wherein MT could improve security and reliability

Anchoring Large Language Models and Digital- Twins in Designing Deeply Personalized mHealth Platforms

Author Name: Muhammad Irfan Khalid, Joe Zhao, Aamir Shahzad, Muhammad Javed, and Mansoor Ahmed

Abstract: This paper presents an automated framework that uses Large Language Models (LLMs) with prompt engineering to generate scalable, measurable and dynamically personalized health goals. By integrating Human Digital Twins (HDTs) the framework leverages behavioral and historical data to continuously refine goals and deliver hyper personalized recommendations. Study results demonstrate that combining LLMs and HDTs overcomes the limitations of static health interventions highlighting the potential of adaptive AI-driven personalization to support sustained behavior change and improved health outcomes

lf-Adaptive Swarm-Driven Big Data Pipelines for Dynamic Workload Optimization in Spark Environments

Author Name: Zulfiqar Ali, Israr Ur Rehman, Kashif Manzer, and Muhammad Abdul Basit Ur Rahim

Abstract: The exponential growth of data-intensive applications has established Apache Spark as a leading platform for large-scale distributed data processing. Despite its widespread adoption, Spark-based big data pipelines typically rely on static configuration parameters that are unable to adapt to dynamic and heterogeneous workload conditions. This limitation often results in performance degradation, increased execution latency, and inefficient utilization of computational resources, particularly in streaming and real-time analytics environments. To address these challenges,
this paper proposes a self-adaptive swarm-driven optimization framework for dynamic workload optimization in Spark environments.
The proposed approach integrates Swarm Intelligence techniques, specifically Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO), with a real-time monitoring and feedback mechanism to automatically tune critical Spark configuration parameters, including executor memory, CPU cores, and data partitioning strategies. The framework continuously analyzes runtime performance metrics and adaptively refines system configurations in response to changing workload patterns.
Extensive experimental evaluation using standard big data benchmarks demonstrates that the proposed framework significantly improves system performance compared to static and rule-based tuning approaches. The results show a 40% reduction in execution latency, an increase in throughput to 14,800 records/s, and an improvement in CPU utilization from 55% to 84%. The main contributions of this work are threefold: (i) the development of a novel self-adaptive swarm-based optimization model for Spark-based big data pipelines, (ii) the integration of real-time feedback driven configuration tuning within distributed processing environments, and (iii) comprehensive empirical validation demonstrating scalability, robustness, and performance gains. The proposed framework provides an

ARQM-LITE: A Small Tool for Requirements Quality Analysis

Author Name: June Porter, Satish Mahadevan Srinivasan

Abstract: Here we present the ARQM-LITE, a lightweight iteration of the Automated Requirements Quality Management (ARQM) tool designed to automate the detection of quality violations in natural language requirements while significantly reducing computational overhead. Building on the original ARQM's BERT-based approach, ARQM-LITE integrates rule-based heuristics through a Soft Match feature, Retrieval-Augmented Generation (RAG) for contextual knowledge, and lighter semantic models. The various components of this model enable span-level semantic scoring and POS tagging with minimal resource demands. When evaluated on several real-world requirement documents and a manually annotated quality dataset, ARQM-LITE achieved faster processing times than ARQM, comparable or superior accuracy on Feasibility and Verifiability, with competitive recall on Singularity. The tool also provides user-friendly PDF and interactive HTML visualizations with explanatory feedback. Results demonstrate that ARQM-LITE successfully balances performance and practicality for requirements engineering practitioners.

FlowAttester: Towards a Runtime Flow Attestation Framework in Programmable NeoCloud Networks

Author Name: Li Wang and Anbang Ruan

Abstract: NeoClouds are emerging as a new generation of cloudplatforms designed to meet the rapidly growing demands of AI workloads.
To support ultra-low latency, massive bandwidth, and _exible networkorchestration, NeoCloud infrastructures rely on Software-De_ned Net-working (SDN) technologies, which provide _ne-grained control on tra_c_ows and can dynamically adjust runtime network behaviors. SDN consists of the control plane, which is in charge of generating and distributing various network _ow rules, and the data plane, which is responsible for receiving and enforcing the _ow rules during runtime.
However, the existing SDN architecture lacks a mechanism to verify whether _ow rules are correctly enforced at the data plane. Consider the control plane delegates _ow rules enforcement to the data plane, the integrity of runtime _ow rules enforcement can be easily compromised by untrusted data plane devices. For example, a malicious user may compromise a data plane device and tamper the tra_c _ow rules.
To address this challenge, we propose Flow Attester, a new runtime attestation framework to verify _ow rules enforcement in programmable NeoCloud networks. We introduce Trusted Computing technologies into the SDN environment to attest runtime _ow rules enforcement of the data plane devices. With our method, the runtime _ow-rule enforcement of a data-plane device can be periodically attested through an attestation protocol, and the attestation results show veri_able evidence of policy-compliant enforcement. We demonstrate the e_ectiveness of our method through preliminary experiment results.

Machine Learning Techniques for Predominant Musical Instrument Recognition in Polyphonic Audio: A Survey of Time-Frequency Representations and Deep Learning Architectures

Author Name: Christian Cimaglia, Muhammad Abid

Abstract: Predominant musical instrument recognition in polyphonic audio is an important task in Music Information Retrieval, with applications in music indexing, transcription, recommendation, and semantic audio analysis. This survey reviews the evolution of machine learning techniques for this task, beginning with handcrafted acoustic features and traditional classifiers, followed by convolutional neural networks operating on time–frequency representations, and more recent transformer, attention-based, involution-based, and modular deep learning architectures.

The review highlights how the field has shifted from manually designed features toward learned representations capable of modeling complex temporal, spectral, and timbral patterns in polyphonic music.

A comparative discussion of representative methods is provided, with emphasis on datasets, input representations, model design, reported performance, and remaining limitations. The survey identifies several open challenges, including efficient long-range temporal modeling, parameterefficient architecture design, inconsistent evaluation protocols, and the trade-off between recognition performance and computational complexity.

These gaps suggest the need for adaptive and efficient time–frequency modeling approaches for future instrument recognition systems.

Affective and Behavioral Computing for Autism Spectrum Disorder Detection: A Survey

Author Name: Jillienne Lapid, Muhammad Abid, and Md Selim

Abstract: Early detection of Autism Spectrum Disorder (ASD) is critical for timely intervention and improved outcomes, yet traditional diagnostic methods remain resource-intensive and subjective. This survey systematically reviews artificial intelligence approaches for automated ASD detection, including facial analysis, behavioral video analysis, eye-tracking and gaze behavior, body pose and movement, selfstimulatory behavior recognition, and brain imaging methods. We examine feature extraction methods ranging from hand-crafted descriptors to deep-learned representations, and classification using machine learning and deep learning architectures. The survey categorizes datasets into pre-training sources, general ASD-focused repositories, and studyspecific clinical collections. We address critical challenges including limited sample sizes, class imbalance, gender bias in neuroimaging, multisite variability, diagnostic validation inconsistencies, and real-world generalization gaps. This comprehensive analysis provides a roadmap for developing robust, clinically relevant AI systems supporting early ASD screening and personalized intervention.

Indirect Feedback Alignment: A Viable Alternative for Deep Neural Networks Backpropagation

Author Name: Jonathan Teck Meng Liaw, Pulin Agrawal, Reza Abdollahi Dezfuli Nezhad, Dheeraj Gollu

Abstract: Backpropagation has been the foundational algorithm for training deep neural networks with great success, driving major advances in artificial intelli-gence, including Large Language Models (LLMs). Despite its effectiveness, it is often considered biologically implausible due to its reliance on symmetric weight transport and precise layer-wise backward error propagation, mechanisms not ob-served in biological neural systems. Recent neuroscientific findings suggest that cortical microcircuits exhibit structured, predominantly forward-directed local signal flow, while broader brain dynamics remain influenced by recurrent feed-back interactions. This motivates exploration of alternative biologically inspired learning mechanisms. This paper introduces Indirect Feedback Alignment (IFA), an uncharted learning framework that Nøkland briefly mentions but has not pre-viously investigated. Unlike backpropagation, IFA avoids strict end-to-end gra-dient transport by introducing an external feedback mechanism that indirectly conveys output error information to early network layers. Inspired by control-theoretic feedback systems such as PID and ADRC, corrective signals guide early representation updates while subsequent layers adapt through forward propaga-tion dynamics. This resembles biological learning, where internal representations are refined through indirect supervisory feedback. Experiments on MNIST and CIFAR using fully connected networks with batch normalization, dropout, and adaptive feedback matrices show test performance generally within ±2% of back-propagation, suggesting IFA as a promising biologically plausible alternative for neural network learning.

RepoWise: A Conversational Framework for Open Source Repository Analysis with Retrieval-Augmented Generation and Structured Analytics

Author Name: Sankalp Kashyap, Arjun Ashok, Nafiz Imtiaz Khan, and Vladimir Filkov

Abstract: Open source software (OSS) stakeholders routinely need answers about project documentation, contributor activity, governance, and community health, yet empirical research documents persistent barriers to accessing this information. We present RepoWise, a conversational framework that enables natural-language interaction with OSS repositories through domain-specialized retrieval. RepoWise unifies semantic search over project documentation and structured analytics over commit and issue data, routed by an LLM-based intent classifier. The tool is publicly hosted with Mistral 7B as the default model, and can also be run locally with any model available through Ollama. To ground the tool's design in documented stakeholder needs, we derived a question taxonomy of 98 questions from 65+ sources, including peer-reviewed studies, industry surveys, and governance frameworks. We evaluate RepoWise on 200 question instances across ten repositories and compare accuracy against DeepWiki and Claude with GitHub MCP. Claude with GitHub MCP achieved the highest quantitative accuracy at 86.7%, followed by RepoWise at 77.5%, while RepoWise requires no paid API access and runs entirely on local hardware. A capability analysis reveals that documentation and analytics questions are well served by current tools, while community health and socio-technical questions lack reliable, verifiable support from current tools. The tool, taxonomy, and evaluation artifacts are publicly available.

An Explainable Neuro-Fuzzy Multimodal Fusion Approach for Lightweight Edge-based Plant Disease Detection in Precision Agriculture

Author Name: Muhammad Abdullah , Tanzila Kehkashan , Domagoj Tuličić , Nikola Ivković , and Adnan Akhunzada

Abstract: Plant diseases pose a significant threat to global food security, causing annual crop losses estimated at 20-40% and demanding intelligent monitoring systems for early detection and intervention. Automated disease classification using deep learning has shown remarkable success on benchmark datasets; however, existing approaches predominantly rely on single-modality RGB imagery and lack the interpretability essential for practical agronomic decision-making. Current fusion strategies fail to adequately model uncertainty and inter-modal dependencies inherent in real-world agricultural environments. This research proposes a novel neuro-fuzzy multimodal fusion architecture that integrates lightweight depthwise separable convolutional networks with adaptive neuro-fuzzy inference systems for edge-deployable plant disease classification. The framework processes RGB, thermal, and multispectral modalities through parallel feature extractors, fusing representations via learnable Gaussian membership functions optimized using Particle Swarm Optimization. Experimental evaluation on the PlantVillage dataset comprising 20,638 images across 15 disease classes demonstrates 99.13% test accuracy with only 950,968 parameters, outperforming comparable lightweight architectures while maintaining a compact 3.63 MB footprint. The proposed approach bridges the gap between black-box deep learning and explainable agricultural decision support, offering a practically deployable solution for resource-constrained edge devices in precision farming applications.

Scalable and Interpretable Text Intelligence for Fake News Detection

Author Name: Imran Ashraf , Tanzila Kehkashan , Roko Milo˘sevi´, Nikola Ivkovi´, and Adnan Akhunzada

Abstract: The dissemination of misinformation has become increasingly severe with the advent of online media, and there is a need for systems that would be able to detect fake news on the internet that would be accurate, transparent, and computationally efficient. These recent deep learning and transformer-based models have been able to provide high-performance predictions, but they are generally resourceintensive and lack interpretability, which might be a disadvantage in resource-limited and transparency-driven contexts. Despite the impressive results of recent deep learning and transformer architectures, their computational requirements and lack of interpretability can limit their use in resource-constrained and transparency-sensitive applications. The aim of this work is to create a light and interpretable Text Classification framework for fake news identification using the use of Term Frequency-Inverse Document Frequency (TF-IDF) feature depiction and Logistic Regression and Support Vector Machine (SVM) classifiers. The proposed model is tested on the Fake and Real News Dataset with 10-fold crossvalidation and hyperparameter tuning. The experimental results illustrate that the accuracy of SVM is 99.57% and the F1 score is 0.9955, while for the Logistic Regression, the accuracy is 98.88%. The results show that well-designed classical ML models can serve as a competitive, scalable, and interpretable benchmark for fake news detection, in the given data set settings. The study also examines the weaknesses of lexical feature representations, such as not having enough semantic understanding and needing external cross-domain validation in future studies.

SA-NET: A Self-Attention Approach for Early and Accurate Skin Cancer Diagnosis Toward Intelligent Dermatology

Author Name: Imran Ashraf , Tanzila Kehkashan , Ivan Mihaljevi´, Nikola Ivkovi´, and Adnan Akhunzada

Abstract: Skin cancer is one of the most dangerous and common diseases and needs to be detected early and classified properly for treatment.

Though conventional machine learning-based models have been widely used for classification, the performance of these models is limited because they require human-designed features. The deep learning techniques achieve very high accuracy, but the important lesion features are not sufficiently extracted, resulting in misclassification. This work aims to design a Self-Attention Network (SA-Net) for better feature extraction for multi-class skin cancer classification. The designed model is designed to be a convolutional model combined with attention mechanisms to focus on important features of the lesions while ignoring background information.

The skin cancer dataset (ISIC) with nine classes is employed, and the model is trained for 25 epochs with a categorical cross-entropy loss function and an Adam optimizer. Experimental performance results in 99.34% accuracy, 98.44% F1-score, 98.74% precision, and 99.28% recall, which are more than the other CNN-based models. It concludes the model's capability in improving the classification performance, further bolstering AI-assisted skin diagnosis.

HERO: Modelling Cost-Energy-Latency Trade-offs in Hybrid Quantum-Classical Cloud Workloads

Author Name: Nirmit Dagli, Prof. Taskin Kocak, Prof. Rongyu Lin, Prof. Kruti Shah, and Prof. Sanjeevkumar Marimekala

Abstract: We present HERO, a modelling-and-measurement software-engineering framework that quantifies the joint cost, energy, latency, and accuracy tradeoffs of running hybrid quantum-classical workloads on today's cloud platforms. Cloud providers offer a uniform Python interface to Central Processing Unit (CPU), Graphics Processing Unit (GPU), and Quantum Processing Unit (QPU) tiers. However, engineers do not have a reproducible method to determine the right tier that a task should target. HERO fills that tooling gap. In this paper, we present case studies specifically designed to exercise opposite ends of the hybrid spectrum. Specifically, we consider a cybersecurity intrusion-detection task on KDD Cup 99 in which a Quantum Support Vector Machine (QSVM) is benchmarked against four strong classical classifiers. The other case we consider is a molecular ground-state estimate of the H2 molecule using the Variational Quantum Eigensolver (VQE) which is benchmarked against exact diagonalization. We validate the framework with two contrasting case studies engineered to exercise opposite ends of the hybrid spectrum. All metric values from each (workload, method) cell are averaged across 30 independent seeds with reported mean and std deviation. The data gives two conflicting results. F1 score = 0.998 at 22 J/task for Random Forest on the cybersecurity dataset and a score of 0.667 for QSVM at around 12,970 times more energy and costing USD 2.1 M/million classification. Regarding molecular simulation, exact diagonalisation is superior at H2 scale but it will hit an O(2^n) memory wall around fifty qubits, after which VQE will still be doable. We create a workload-aware tier-allocation algorithm with suitable thresholds and 2026 cloud-cost estimates. SE community may find value in the simulator we release as open-source artifact.

Engineering Reliable Image-Forgery Localization Systems: A Multi-Dataset Study of Topology-Aware Vision Transfor

Author Name: Lavanya Mandava, Rohit Biswa Karma, and Shanmukh

Abstract: Image-forgery localization systems aim to identify manipulated pixels, yet their reliability as deployable ML software remains difficult to assess. A method can perform well on one benchmark while failing under dataset shift, target-domain adaptation, threshold changes, or authentic-image conditions. This paper presents a multi-dataset empirical evaluation of image-forgery localization as an ML software reliability problem, using a topology-aware Vision Transformer pipeline as the studied system. The pipeline combines ViT patch features, dense pixel refinement, and graph-based regularization over learned patch relationships.
We evaluate the system on CASIA, COVERAGE, and CoMo-FoD under in-domain, zero-shot, fine-tuned, joint-training, adaptation, and robustness settings, and we compare it with TruFor and MVSS-Net on a validated fixed published-baseline subset. The results show that graph regularization and pixel refinement improve the matched CASIA model, but CASIA-only training transfers poorly to external copy-move datasets. Joint CASIA+CoMoFoD training improves target-domain performance and provides a stronger initialization for COVERAGE adaptation, but authentic-image false positives remain a major reliability risk.
The study shows that image-forgery localization should be evaluated not only by tampered-image IoU or F1, but also by multi-dataset transfer, adaptation behavior, threshold diagnostics, prediction/mask sanity checks, and authentic-image false-positive metrics.

The Effectiveness of Feedback Design for Improving LLM-based Code Generation in Quantum Optimization Algorithms

Author Name: Venkata Sai Prathyush Turaga, Sonali Uttam Singh, and Akbar Siami Namin

Abstract: Hybrid quantum-classical programs often fail silently. A Quantum Approximate Optimization Algorithm (QAOA) MaxCut program can execute without exhibiting any faulty behavior and errors yet return a wrong answer because bugs can hide in independent components such as Hamiltonian construction, bitstring decoding, and cut-value calculation.

Standard simulator feedback reports only the final outcome, such as a low approximation ratio or a wrong answer, without localizing the cause of failure or identifying which component failed. This paper investigated whether providing an LLM with component-level diagnostics improves its ability to repair, and thus enhance its own generated code. We conduct controlled experiments across 2,160 trials spanning six LLMs (devstral:24b, granite4.1:30b, mistral-small3.2:24b, gemma4:26b, gpt-oss:20b, and nemotron3:33b), six graph families (Erdős–Rényi, random 3-regular, complete, cycle, path, and ladder), and five graph sizes ($n \in \{4, 6, 8, 10, 12\}$). The research work compares three feedback designs: (1) Outcome-Level Feedback (OLF) (Arm A), (2) Component-Level Diagnostics (CLD) (Arm B), and (3) Diagnostics with Concrete Counterexamples (DCC) (Arm C). According to the result, Component level diagnostic feedback (CLD) raises the final pass rate by 6.8 percentage points (pp) overall and by 30 percentage points (pp) for 'gpt-oss:20b', which is the only model to exceed 90% (62.5% to 92.5%). More detailed feedback costs fewer tokens, not more, because precise diagnostics cause programs to converge faster. Counterexample samples add a further 2.3 percentage points (pp) on average and help standard coding models more than reasoning models. These results show that the structure of feedback, not only its presence, determines how effectively an LLM repairs generated quantum programs.

Eye-RIS: Design and Evaluation of a Low-Latency Voice-Driven AR System for Real-Time Clinical Document Analysis on Smart Glasses

Author Name: Robert John Luther Bennethum IV, Zainab Elahi Qazi, Ozge Selin Ak, Logan Swartz, Griffin Garman, Isiah Milien, Matthew Pragel1, Paul Kocis, Sultan Al Nahian, and Sayed Mohsin Reza

Abstract: The increasing adoption of smart glasses in clinical environments necessitates efficient, hands-free solutions for accessing and analyzing medical information in real time. This paper presents Eye-Reality Interactive System (Eye-RIS), a low-latency, voice-driven augmented reality (AR) system designed to enable real-time clinical document analysis on smart glasses. Eye-RIS integrates natural language voice interaction, large language model (LLM)-based document summarization, and keyword-based retrieval into a seamless heads-up display (HUD), thereby allowing healthcare practitioners to maintain focus on patient care while accessing critical information. The system is built on a scalable, cloudnative architecture that uses AWS services and a high-speed inference engine to achieve quick response times. A dual-stage processing pipeline combines OCR-based text extraction with semantic search to support both structured and unstructured clinical documents in PDF format.

To ensure usability in clinical settings, the platform adopts a lightweight web-based interface compatible with available AR devices, enabling hardware portability and rapid deployment. We evaluated Eye-RIS through a combination of system-level testing, performance benchmarking, and user-centered usability studies. Experimental evaluation demonstrates that Eye-RIS achieves significant performance improvements, including substantial gains in processing speed and responsiveness, while maintaining reliable transcription accuracy. User feedback further indicates enhanced accessibility and reduced cognitive load during document interaction tasks. Our results suggest that voice-driven AR systems such as Eye-RIS can substantially improve clinical workflows by minimizing device switching and enabling real-time, context-aware document analysis.

This work highlights the potential of combining AR, AI, and cloud computing to advance next-generation healthcare interfaces and provides a foundation for future research in low-latency intelligent wearable systems.

VALS: A Model-Agnostic Virtual Reality Framework for Real-Time Annotation of 3D Anatomical Models

Author Name: Pranav Balachander, Lasya Madhuri Gundapaneni, Yamini Satish and Khush Mistry, Evan Goldman, and Sayed Mohsin R

Abstract: Traditional anatomy education relies on two dimensional diagrams and static physical models, which fundamentally limit ability to develop an intuitive understanding of complex three-dimensional spatial relationships. This paper presents VALS (Virtual Reality Anatomy Labelling System), a model-agnostic virtual reality framework developed for the Meta Quest 3 headset that enables anatomy instructors and students
to load, manipulate, and annotate three-dimensional anatomical models in an immersive environment. VALS supports runtime loading of arbitrary .OBJ model files directly from headset-local storage, real-time model manipulation via Meta Quest controller input, and an interactive pin-based labeling system that incorporates a Trie-based search structure to support low-latency anatomical term lookup during text input. Label data is persisted locally through a lightweight JSON serialization system, enabling seamless session continuity without the overhead of re-exporting mesh geometry. The framework is implemented in Unity 6 using C# with OpenXR and the Meta XR SDK, allowing deployment on standalone consumer hardware without reliance on cloud infrastructure. Performance evaluation on the Meta Quest 3 demonstrated consistent rendering at 72 FPS under standard usage conditions, near-instantaneous model loading (under 1 second), and label placement latency below 0.5 ms. Trie-based anatomical term retrieval completed in under 10 ms. A structured usability study confirmed that the core instructor workflow: loading a model, placing labeled pins, and saving a session can be completed within minutes by first-time users. VALS addresses a clear gap in accessible, immersive anatomy tooling by providing a lightweight and extensible VR framework for real-time annotation of user-defined anatomical data.

Phishing Attack Detection using Genetic Algorithms and Artificial Neural Networks

Author Name: Ali Al-Mehaizaa(B), Mohammed Al-Mansoori, Mohammed Al-Mohannadi, Abdullah Al-Ghushami, and Adnan Akhunzada

Abstract: Phishing remains one of the most prevalent cyberattack vectors targeting individuals and organizations through fraudulent URLs distributed across email, SMS, social media, and instant messaging channels.Static rule-based and signature-based detection mechanisms struggle to keep pace with the rapidly evolving tactics employed by attackers. This paper presents a hybrid phishing detection system that integrates Genetic Algorithms (GAs) for feature selection with an Artificial Neural Network (ANN) for classification. The GA evolves an optimal subset of URL features from a corpus of 87 candidate indicators spanning URL structure, domain registration, content characteristics, and third-party verification markers, drawn from the publicly available Hannousse & Yahiouche phishing-website benchmark dataset (11,430 URLs), which is itself compiled from PhishTank and from a curated set of legitimate URLs. The selected features (45 of 87) are then fed to a feedforward ANN with two 64-unit ReLU hidden layers and a sigmoid output, trained with the Adam optimizer and binary cross-entropy loss. Feature selection and model evaluation are separated by a stratified 80/10/10 train/validation/test split: the genetic algorithm uses only the training and validation partitions to select features, and all final metrics are reported on a held-out test set that is never used during selection, eliminating validation-set leakage. Averaged over five random seeds, the proposed system achieves accuracy 0.938 □} 0.006, precision 0.938 □} 0.006, recall 0.938 □} 0.006, and F1-score 0.938 □} 0.006 (macro-averaged across both classes) on the held-out 1,143-sample test partition. The genetic algorithm is benchmarked against both random feature subsets and a mutual-information feature-selection baseline at matched cardinality: it outperforms random selection (accuracy 0.859) by 7.9 percentage points and the mutual-information baseline (accuracy 0.827) by 11.0 points, confirming that the evolutionary search identifies a meaningfully better feature combination than either chance or a standard univariate selector.
The principal limitations are dependence on the diversity and recency of the training dataset and the absence of full k-fold cross-validation;

A Hybrid-Simulation LoRaWAN Honeypot for Detecting RF and Protocol Attacks against IoT Networks: A Cyber-Electromagnetic Activities Perspective

Author Name: Ali Al-Mehaizaa(B), Gabriele Oligeri, Faisal Al-Bishri, and Adnan Akhunzada

Abstract: The proliferation of Low-Power Wide-Area Network (LPWAN) deployments — and LoRaWAN in particular — has expanded the cyber-electromagnetic attack surface of modern IoT systems. While LoRaWAN's protocol-layer security has been extensively analysed, defensive instrumentation that targets the radio side has received comparatively little attention. This paper presents the design, implementation, and evaluation of a radio-side LoRaWAN honeypot system architected from a Cyber-Electromagnetic Activities (CEMA) perspective. The honeypot is positioned as an Electronic Support capability that informs Electronic Protection, Cyberspace Operations, and Spectrum Management Operations. It comprises a swappable RF front-end, a frame-level traffic synthesis layer, an IQ-level physical-layer simulator, a multi-detector inspection pipeline covering five protocol-layer attacks, and a Random-Forest-based classifier for jamming-style physical-layer attacks. Six attacks drawn from the LoRaWAN threat literature are evaluated: replay,frame-counter manipulation, join-procedure replay, MIC-failure flooding, rogue-device presence, and reactive jamming. The evaluation is organized in two stages. Stage 1 runs the detection pipeline against a controlled simulation in which both legitimate traffic and attack instances are generated by parameterisable reference implementations; this stage yields a true-positive rate of 1.000 on four of the five protocol attacks and 0.930 □} 0.051 on frame-counter manipulation across five 2-hour trials, with zero false positives over a baseline of 1,727 legitimate frames per detector. For the five-class jamming classifier, the Random-Forest model perfectly classifies the three structured jammer types (continuous-wave, pulsed, and sweep; per-class F1 = 1.000) but distinguishes reactive jamming from the no-jamming baseline only at chance (per-class F1 ≈ 0.50), yielding an aggregate accuracy and macro F1 of 0.800; we trace this limitation to the use of episode-global features and identify burst-aware extraction as the remedy. Stage 2 validates the system against real captured

Generative AI, Agentic AI, MCP, and RAG for Improving Developer Productivity

Author Name: Rakesh Kumar Bidanagere Nagaraju

Abstract: Large-scale enterprise software development is increasingly con-strained by context deficits rather than engineering capability gaps, as develop-ers navigating thousands of interdependent services spend disproportionate time on context recovery rather than active problem-solving. A converged platform architecture combining Generative AI, Retrieval-Augmented Generation, Mod-el Context Protocol, and Agentic AI workflows addresses this challenge by transitioning developer tooling from passive information retrieval to active, rea-soning-driven assistance. Retrieval-Augmented Generation grounds models with internal knowledge (service ownership, API contracts, and incident histo-ry). Combining data with the Model Context Protocol, which adds clearly de-fined context boundaries, makes hallucinations less likely and enables produc-tion-quality results. Agentic AI orchestration handles multi-hop diagnostics, making and validating hypotheses against the complex, entangled workings of an enterprise ecosystem where interventions are treated as hypotheses that must be validated or falsified. Integrated architectures radically reduce context acqui-sition time and toil and shift institutional knowledge into engineering, whether IDEs or CI/CD pipelines. The foundational insight across all deployment con-texts is that AI effectiveness is governed by context engineering quality rather than model sophistication alone and that organizations treating retrieval pipe-lines and context governance as first-class infrastructure build compounding and durable advantages in engineering velocity.

Human-Grounded AI-Assisted Software Engineering Education: An Empirical Study of Learning Outcomes

Author Name: Junaid Shuja and Abir Saha

Abstract: Software engineering has been an integral part of the AI revolution as several clients facing tools are being developed leveraging the power of AI agents and large language models. More software developers and practitioners are utilizing AI tools to write code, draw system models, generate test cases, and specify software requirements. As a result, the teaching of software engineering in higher education needs to evolve and integrate the AI tools in the classroom. Software engineering students are taught the life-cycle of software development which starts at project initiation and leads towards phases of requirements engineering, software modeling, design, architecture, testing, and deployment. Each of these phases embeds AI in practice. Therefore, it is necessary that the same AI tools are integrated in teaching and learning software engineering to educate market-ready graduates. We introduced a Human-Grounded AI-Assisted approach to software engineering education. We conducted an empirical study to answer the following research questions.
RQ1: Did students achieve better course learning outcomes because of AI-integration? RQ2: How did human-generated requirements, UML diagrams, and test cases compare with AI-generated artifacts in terms of correctness.We integrated AI tools across two semesters of a software engineering course, requiring students to first produce UML and design artifacts manually before recreating them using AI tools. Submissions from 13 groups in Fall and 18 groups in Spring semester were evaluated using instructor rubrics and Claude-based UML validation. The empirical analysis shows that the Human-Grounded AI-Assisted approach resulted in better learning outcomes. Moreover, AI-generated artifacts consistently outperformed human-generated artifacts in technical correctness across both semesters.

An Empirical Investigation of Soft Skills and Corresponding Challenges in International Software Engineering Education

Author Name: Sajid Ibrahim Hashmi, Jouni Markkula

Abstract: Joint software development work requires the application of technical skills and professional practices. Soft skills, such as teamwork, remain essential to succeeding in a professional software development environment. The phenomenon is reflected by group work in SE (software engineering) education, where students can learn and practice teamwork-related soft skills. Therefore, SE education requires students to practice those skills so that they can actively embark on their professional careers in the software industry.

In this research study, empirical findings from two focus groups we conducted on group work are presented, with emphasis on the soft skills necessary for teamwork in international software engineering education, specifically student collaboration in higher education. There are several soft skills students should learn as they practice teamwork in their course-related assignments and exercises, as well as the challenges that come with it in international settings. The findings serve as a guideline to university teachers and pedagogy experts to plan student group work more effectively.

AI Turning Static Exports into Actionable Health Intelligence

Author Name: Sagar Rao, Atif Farid Mohammad, Paige Coleman, Sarthak Bhatt

Abstract: Electronic Health Information (EHI) exports have emerged as a regulatory imperative in U.S. healthcare policy, yet widespread usability failures have undermined their clinical and patient-facing value. A recent independent review assigned a median grade of D to the export experience across 265 certified Electronic Health Record systems (Mandel, 2026), demonstrating that data availability alone does not produce actionable intelligence. This paper presents the Healable platform, an AI-powered orchestration framework that transforms fragmented Single-Patient EHI exports into longitudinal, explainable, and role-aware patient records. The system employs a governed agentic AI architecture combining retrieval-augmented generation (RAG) with Context-Augmented Generation (CAG; Mohammad, 2026) to ground every output in the patient's own health data. Specialized agents handle clinical summarization, medication reconciliation, care-transition analysis, population risk stratification, patient engagement, social determinants of health (SDOH) detection (Mohammad et al., 2021; Bearse et al., 2020b), and payer coordination, within a governance framework informed by responsible-AI principles (Coleman & Mohammad, 2025; Kalva & Mohammad, 2025). Data normalization employs graph-theoretic harmonization (Mohammad et al., 2020a). A reference deployment against a synthetic multi-provider record demonstrates the system's capacity to compress a 6,200-resource FHIR bundle into role-specific, cited, actionable outputs while preserving full source attribution and human oversight.

Self-Healing Cloud Infrastructure: Integrating Multi-Agent Reinforcement Learning with Policy-Driven Autonomy for Resilient Enterprise Operations

Author Name: Balasubramanian Bava Jagannathan

Abstract. Enterprise cloud platforms have become the operational back-bone of financial, insurance, and technology industries, yet the incident response frameworks governing them remain predominantly reactive—dependent on threshold-based alerting, human-in-the-loop escalation chains, and static remediation runbooks that are structurally ill-suited to the ve-locity and complexity of modern failure cascades. Artificial Intelligence for IT Operations (AIOps) has advanced anomaly detection and root-cause inference but stops short of autonomous remediation, leaving mean time to recovery (MTTR) as a persistent operational liability. This pa-per proposes a Self-Healing Cloud Architecture that embeds multi-agent reinforcement learning (MARL) within a policy-governed control loop, enabling fully autonomous fault detection, lassification, and remedia-tion while satisfying enterprise governance constraints. The architecture comprises four coordinated layers: a telemetry and fault detection layer, a policy and safety enforcement layer, a multi-agent decision layer host-ing specialized agents for fault classification, remediation planning, and capacity management, and a Kubernetes-native execution and orches-tration layer. The remediation optimization problem is formalized as a Constrained Markov Decision Process (CMDP) with a multi-objective reward function balancing recovery speed, availability service-level objec-tive (SLO) compliance, and remediation resource efficiency. Simulation-based evaluation across node failure, traffic surge, and cascading degra-dation scenarios shows that the proposed system reduces median MTTR by 56%, improves availability SLO compliance by up to 4.9 percent-age points, and reduces remediation over-provisioning by 18 percent-age points compared to rule-based auto-recovery and AIOps-assisted baselines, while maintaining 100% governance policy compliance. These results establish a credible and deployable path toward genuinely self-healing enterprise cloud platforms.

An Empirical Study of How Publication Context Relates to Repository Maintenance and Community Adoption in AI Research Repositories

Author Name: Sang Vo, Caijing Yang, Xuan Nguyen, Boshuai Ye, Arif Ali Khan

Abstract: The growing practice of releasing code alongside AI papers has led to numerous accessible research repositories. While valuable as reusable academic assets, their post-publication maintenance and community adoption remain insufficiently understood. This study investigates whether publication context is associated with repository maintenance and adoption. Using a PapersWithCode dataset of 576 repositories with 24,953 issues, we analyze the relationship between publication venue, year, and framework selection on several indicators. Specifically, maintenance is evaluated via issue closure and labelling rates, while community adoption is measured by forks, watchers, and contributor counts. Our findings show modest but statistically detectable effects of publication context on repository outcomes. Overall, publication-related factors alone cannot fully explain differences in maintenance and adoption. Additionally, our findings show the gap between merely publishing code associated with a paper and maintaining it as a long-term open-source project.

HERO: Modelling Cost-Energy-Latency Trade-offs in Hybrid Quantum-Classical Cloud Workloads

Nirmit Dagli, Prof. Taskin Kocak, Prof. Rongyu Lin, Prof. Kruti Shah, and Prof. Sanjeevkumar Marimekala

Abstract: We present HERO, a modelling-and-measurement software-engineering framework that quantifies the joint cost, energy, latency, and accuracy tradeoffs of running hybrid quantum-classical workloads on today's cloud platforms. Cloud providers offer a uniform Python interface to Central Processing Unit (CPU), Graphics Processing Unit (GPU), and Quantum Processing Unit (QPU) tiers. However, engineers do not have a reproducible method to determine the right tier that a task should target. HERO fills that tooling gap. In this paper, we present case studies specifically designed to exercise opposite ends of the hybrid spectrum. Specifically, we consider a cybersecurity intrusion-detection task on KDD Cup 99 in which a Quantum Support Vector Machine (QSVM) is benchmarked against four strong classical classifiers. The other case we consider is a molecular ground-state estimate of the H2 molecule using the Variational Quantum Eigensolver (VQE) which is benchmarked against exact diagonalization. We validate the framework with two contrasting case studies engineered to exercise opposite ends of the hybrid spectrum. All metric values from each (workload, method) cell are averaged across 30 independent seeds with reported mean and std deviation. The data gives two conflicting results. F1 score = 0.998 at 22 J/task for Random Forest on the cybersecurity dataset and a score of 0.667 for QSVM at around 12,970 times more energy and costing USD 2.1 M/million classification. Regarding molecular simulation, exact diagonalisation is superior at H2 scale but it will hit an O(2^n) memory wall around fifty qubits, after which VQE will still be doable. We create a workload-aware tier-allocation algorithm with suitable thresholds and 2026 cloud-cost estimates. SE community may find value in the simulator we release as open-source artifact.

Idea and Vision

Development of a Child-Friendly Constraint-Induced Movement Cast Using 3D Scanning and 3D Printing

Author Name: Remington Orange, Bibiana M. Steckel Christopher R. Shelton

Abstract - Background: Constraint-Induced Movement Therapy (CIMT) is an evidence-based intervention for improving upper-limb function in individuals with hemiparesis. However, outcomes may vary depending on the type of constraint device used. Splint-based constraints have been associated with more promising outcomes than mitts/gloves or slings, but available pre-made and custom thermoplastic splints may be difficult to use in pediatric populations due to fit-related issues, discomfort, and fabrication challenges. Advances in digital fabrication, including 3D scanning and 3D printing, may offer new opportunities to redesign CIMT constraint devices in a more comfortable, safe, accessible, and child-friendly format.
Objective: To describe the development of a child-friendly, 3D-scanned and 3D-printed CIMT constraint device intended to support upper-limb motor rehabilitation in children with hemiparesis.
Methods: A customized CIMT constraint prototype was developed using 3D scanning, digital modeling, 3D printing, and Velcro-based assembly. The prototype was developed using an internal adult reference scan and was not tested with pediatric participants. A researcher's upper limb was scanned, modeled in Blender, divided into three printable sections, sliced in Prusa Slicer, and printed using a Prusa MK4S printer.
Results: The workflow resulted in a three-section 3D-printed constraint prototype covering the hand, forearm, elbow, and upper arm. The device included an open lattice structure intended to reduce weight and support airflow, with Velcro fastening used for adjustability and positioning. Because the prototype was developed from an adult reference scan, the present results are limited to technical feasibility and prototype fabrication rather than pediatric fit, comfort, usability, safety, or clinical effectiveness.
Conclusion: The prototype supports the technical feasibility of using 3D scanning and 3D printing to create a personalized, breathable, and adjustable CIMT constraint device. However, the

device has not yet been evaluated with children. Future pediatric testing is needed to assess comfort, safety, usability, acceptance, and suitability for CIMT-related rehabilitation contexts.

Intrinsic Training and Compounding Accuracy: Designing AI-Assisted SAS-to-R Migration in Clinical Trials

Author Name: David Ju and Nathalia Nascimento

Abstract: Statistical programming teams in clinical trials face growing pressure to migrate from SAS to R while maintaining regulatory traceability and work-force skill continuity. AI-assisted code-translation tools are typically evaluated using a single accuracy score from a single benchmark run, a framing that is in-adequate for regulated clinical settings where reviewers must reconstruct the lin-eage of every computed value and teams must transition skills alongside tools. We propose a design vision in which AI-assisted workflows are evaluated as sys-tems of compounding accuracy inseparably linked to hands-on training. In this vision, every reviewer intervention serves two purposes: it grows the agent's cor-pus, improving subsequent translations, and it produces an inspectable record that domain reviewers can audit. Each intervention is captured as a structured entry that the agent can retrieve during the next translation and that a domain reviewer can inspect. Drawing on clinical-trials industry experience, we articulate three design principles, namely training as a deliverable, capture-and-replay accuracy, and reviewability over autonomy, each paired with a testable hypothesis to guide future empirical work. We instantiate the principles in an illustrative worked ex-ample using a Microsoft 365 Copilot agent, a curated reference document, and a PROC COMPARE-oriented validation workflow. The worked example demon-strates that the principles can be instantiated in a concrete workflow and supports the feasibility of the testable predictions. The framework is accessible to any or-ganization with an existing SAS environment and a configured LLM agent and provides a research agenda for future empirical validation.

AI Danger Perception, Ethical Use, and Professional Regulation: A Framework for Algorithmic Alignment and Risk Management

Author Name: Intisar Rizwan I Haque and Atif Farid Mohammad

Abstract: As artificial intelligence (AI) systems transition from passive, tokengenerationtools to autonomous, decision-making agents in safety-critical and licensed domains, aligning machine behavior with human safety remains a critical challenge. While biological systems possess an intrinsic perception of danger grounded in autopoietic self-preservation, machine learning systems lack this existential anchor, operating purely to optimize extrinsic, mathematical proxy rewards.

This paper examines this misalignment through a unified, interdisciplinary framework, identifying three key technical deficits in contemporary architectures: (1) the absence of autopoietic self-preservation, which induces severe reward hacking and alignment faking; (2) stepwise temporal myopia, which undermines long-horizon planning and decision-making; and (3) majoritarian preference collapse, which erases pluralistic cultural and demographic values. To bridge the gap between these computational limitations and real-world safety, we advocate adopting a structured, six-step professional clinical licensure and credentialing framework that mirrors human medical credentialing. By synthesizing technical machine learning evaluations with professional regulatory models, we provide a multi-layered roadmap for robust AI risk management and systematic socio-technical governance

Short Research Paper

Comparative Study of Security Testing Tools for Distributed Web Applications

Author Name: Bouchaib Falah, Grace Julius, Onyedikachi Kanu, and Chinyere Offor

Abstract. Distributed web applications run across multiple servers and networks and are now widely adopted, which has made securing them a serious challenge. These applications face increased exposure to vulnerabilities such as SQL injection, cross-site scripting, authentication flaws, and misconfigurations, yet selecting an appropriate security testing tool remains difficult due to varying capabilities, accuracy, integration support, and scalability. This paper presents a comparative study of six leading security testing tools: SonarQube, Checkmarx, Acunetix, Postman, Contrast Security, and Dependabot. The goal is to evaluate their effectiveness in identifying and mitigating risks in distributed environments.
We employ a hybrid approach that combines an extensive literature review with hands-on exploratory testing on deliberately vulnerable applications, assessing each tool against detection capability, performance and scalability, coverage, automation, and usability. No single tool addresses every vulnerability: static tools excel at early code-level detection, dynamic and API tools surface runtime and endpoint issues, interactive analysis offers context-aware accuracy, and dependency scanning secures third-party components. The study offers practitioners a criteria-based framework for aligning tool selection with organizational needs such as scalability, CI/CD integration, and ease of use.

CASL: A Software Framework for Gesture-Driven, Context-Aware Layout Design in Virtual Reality

Author Name: Annika Brown, Katie Ho, and Trudi Di Qi

Abstract: Spatial design tasks, yet engaging such systems remains challenging when user input is continuous, imprecise, and constrained by domain-specific design rules. This paper presents the Context-Aware Spatial Layout (CASL) framework for gesture-driven, constraint-aware layout design in VR. CASL combines global layout initialization with localized semantic snapping, interpreting hand gestures as coarse spatial intent and resolving them into nearby, design-compliant configurations. We demonstrate CASL in an interior layout design testbed and evaluate its computational performance across global organization and local gesture-guided refinement scenarios. Results show consistent reductions in layout energy cost and improved refinement performance compared with standard grid-based snapping. CASL offers a practical architecture for balancing embodied user control with domain-aware computational assistance in human-interactive VR applications.

Closing the Inconsistency Trap: Unified Cross-Layer Deception via Adversarial LLM Transformation

Author Name: Osamah Mansoor, Adeel Anjum, Madiha Haider Syed, Dr. Muhammad Talha Gul, and Jie Zhao

Abstract: Remote operating system (OS) fingerprinting is the foundational stage of the cyberattack lifecycle, enabling attackers to map infrastructure vulnerabilities across network layers. Although prior work such as SoFi [1] has demonstrated effective perturbation of Layer 4 (Transport) headers, these approaches remain vulnerable to what we term the Inconsistency Trap: spoofed TCP/IP signatures that contradict raw Layer 7 (Application) service banners, providing a high-confidence forensic signal that reveals the system's true identity. This paper proposes a Unified Semantic Fingerprint Spoofing framework that resolves this architectural gap by deploying a Large Language Model (LLM) as a stateful Deception Controller. The controller optimizes network-layer obfuscation with real-time application-layer identity synthesis using a Weighted Projected Gradient Descent (W-PGD) algorithm guided by Jacobian-based Saliency Mapping.

We further demonstrate that the same adversarial principles extend to automated LLM red teaming through Intent Spoofing, wherein a DL-on-DL surrogate attack framework bypasses LLM safety guardrails. Empirical evaluation across 50 trials demonstrates a BERTScore of 0.982, a protocol compliance rate of 99.8%, and reduction of professional-grade scanner detection (Nmap, ZGrab2, p0f) from 98% to 0.5% ± 0.08%. Intent Spoofing achieves a 98% Attack Success Rate (ASR) against DeepSeek and 90% against GPT-4 while preserving semantic coherence.

VirNetMTD: Programmable Moving Target Defense for Virtualized Networks

Author Name: Thu Dang and Li Wang

Abstract: Cloud and virtualization technologies have fundamentally transformed modern computing infrastructures. Increasingly, applications, services, and large-scale computing workloads are deployed in virtualized environments to benefit from scalability, flexibility, and operational efficiency. However, while virtualization provides significant advantages, it also reshapes the attack surface of modern applications and services, creating new opportunities for large-scale reconnaissance, lateral movement, and adaptive cyber attacks. Meanwhile, recent advances in Large Language Models (LLMs) and AI-driven offensive techniques have significantly reduced the cost and complexity of cyber attacks, enabling attackers to automate vulnerability discovery, network reconnaissance, exploit generation, and attack planning at unprecedented scale and speed.
Moving Target Defense (MTD) has emerged as a promising defense paradigm that aims to break the information asymmetry between attackers and defenders by continuously mutating system properties and invalidating attacker knowledge. However, deploying continuous MTD mutation strategies in conventional infrastructures has historically been challenging due to their static and inflexible nature. The emergence of virtualization, software-defined networking (SDN), and programmable infrastructures now makes continuous and coordinated network mutation practically feasible. In this paper, we present VirNetMTD, a programmable Moving Target Defense framework for virtualized network environments. VirNetMTD leverages virtualization and SDN technologies to enable dynamic and coordinated network mutation at runtime.
Our framework supports multiple MTD strategies, including topology obfuscation, traffic manipulation, traffic redirection, and infrastructure diversification. We implement a prototype of VirNetMTD in a virtualized SDN environment and demonstrate its effectiveness in disrupting network reconnaissance and increasing attack uncertainty. Our results show that virtualization provides a

practical foundation for scalable, adaptive, and programmable Moving Target Defense in modern cloud infrastructures.

Student Research Competition

Collaborative Generation of Structural Bug Signatures via Privacy-Preserving Federated Diffusion Models

Author Name: YueMing Zhang, Minh Tung Le, Aidan Tristen R. Angel, and Muhammad Abdul Basit Ur Rahim

Abstract: Data-driven software defect detection relies heavily on vast, diverse datasets of vulnerable code. However, organizations are often legally prohibited from sharing proprietary source code, leading to localized data scarcity. While Federated Generative Adversarial Networks (Fed-GANs) have been proposed to collaboratively synthesize training data across cross-device scenarios without sharing raw code, they frequently suffer from mode collapse, generating redundant test cases that fail to explore diverse software execution paths. In this paper, we propose a Federated Diffusion Model equipped with a Semantic Quality Filter to collaboratively generate structural code signatures.

By transforming C/C++ source code into 2D structural bytematrices, our model learns to reverse-diffuse structural bug patterns collaboratively. Our evaluation on the CodeXGLUE defect dataset demonstrates that our approach completely avoids mode collapse, maintaining high structural diversity ($L2 \approx 11.0$) across 100 communication rounds even under heavy privacy noise ($\epsilon = 8.0$). Furthermore, when used for test-suite augmentation in a low-resource regime, our synthetic signatures act as a data lifeline, improving downstream bug classifier accuracy by +19.1%.

Empirical Investigation of Cyber Attack Detection

Author Name: Joshua Partridge, Alex Wood, Ghulam Mudassir

Abstract: Ransomware and zero-day attacks are an escalating threat to enterprise networks, and signature-based intrusion detection systems cannot keep up with how quickly attack patterns change. The UGRansome dataset captures cyclostationary network-flow records, financial-impact metrics, and labelled attack signatures, which makes it useful for empirical study of zero-day detection. This work classifies UGRansome network flows into three categories: Anomaly, Signature, and Synthetic Signature, where the Synthetic Signature class is a stand-in for simulated zero-day attacks. The study first examines the distributions, pairwise correlations, and variance inflation factors of the continuous attributes, then trains Logistic Regression and Linear SVM models as linear references. Both linear models plateau at roughly 88% accuracy, and their residual errors concentrate on the Anomaly class, whose recall stays below 83% for both models; this points to class boundaries the linear models cannot represent and motivates a non-linear Random Forest classifier. On a stratified 80/20 split of the 149,043 records the Random Forest reaches 99.29% accuracy and 99.29% weighted F1, with all transforms fitted on the training split only. Feature-importance analysis ranks the financial-impact attributes (USD, BTC) first, yet an ablation shows their removal costs at most 0.44 accuracy points, so the financial signal is the strongest individual predictor but is largely recoverable from the remaining flow attributes. Because a random split can place records from the same ransomware families and addresses in both training and testing, the model is additionally evaluated with five-fold cross-validation, a family-based holdout, a time-based holdout, and a leakage-aware feature set that removes identifier attributes. Cross-validation reproduces the point result (99.30% ± 0.02 accuracy); a family-based holdout over four unseen ransomware families reaches 99.42% weighted F1; the leakage-aware feature set retains 97.78%; and the time-based holdout, the most conservative protocol, retains 94.99%. The reported performance therefore does not rest on memorised identifiers; the time-based figure is the practical lower bound for generalisation to future traffic.

Biologically Plausible Learning in Spiking Neural Networks Using Indirect Feedback Alignment: A study of the implementation of Feedback Alignment methods into Spiking Neural Networks

Author Name: Joshua Nallaraja and Jonathan Liaw

Abstract: Backpropagation (BP) has been the backbone for training deep neural net-works; however, its biological implausibility, particularly the requirement for symmetric weight transport between forward and backward passes, inhibits its implementation in biologically attuned fields. In contrast, an existing biological alternative to the traditional artificial neural network is the recently introduced spiking neural network. By implementing feedback alignment methods of weight propagation, a more biologically plausible solution to the biological gap can be created. This paper presents a systematic implementation of feedback alignment methods into the spiking neural network framework. For this paper, three exper-iments are performed with models of varying SNN builds: a vanilla implementa-tion with backpropagation and both a two layer and three-layer version of the indirect feedback alignment. Identical network architectures, including SNN dy-namics, LIF neuron parameters, Poisson spike encoding and surrogate gradients, are maintained amongst all models for the fairest comparison. The vanilla feed-back alignment model is biologically questionable at best and was therefore not included in this research. All algorithms are evaluated on the MNIST dataset, a handwritten digit recognition benchmark, trained across a comprehensive sweep of five optimizers and six activation functions. As expected, backpropagation achieves the highest accuracy, but it is challenged in several areas by the feed-back alignment models. Our results are cautiously promising, with numbers near-ing the performance of backpropagation.

Enhancing Requirement Analysis with Hybrid AI-Based Question Answering Systems

Author Name: Yongyan Liang, Lihua Lan, Muhammad Abdul Basit Ur Rahim*

Abstract: Requirement analysis is a critical phase in software engineering, but it is often affected by ambiguity, incompleteness, and inconsistency when Software Requirements Specifications (SRS) are written in natural language. Existing AI-based question answering systems improve requirement analysis but remain limited by the effectiveness of their retrieval and ranking components. This research enhances the QAssist framework by introducing BGE embeddings, a BGE cross-encoder reranker, and a hybrid retrieval strategy that combines lexical and semantic methods.

The proposed approach is evaluated using the REQuestA dataset together with GTv6 and DexV2. Results demonstrate improved recall and ranking quality for domain-specific queries while introducing a trade-off in execution time. These findings highlight the importance of semantic retrieval in requirement analysis and demonstrate the potential of AI-based retrieval and question answering systems.

An Exploratory Machine Learning Approach for Mental-Health Related Text Classification

Author Name: Brandon Breindel and Muhammad Talha Gul

Abstract: Machine learning methods may support preliminary screening of mental health related text for more accurate diagnosis in clinical settings. How-ever, social media as well as survey-based approaches require careful interpreta-tion. This study explores whether machine-learning models can utilize social media posts to classify data based on survey results. Reddit posts were collected from four mental health related communities: Autism, Bipolar, OCD, and Schiz-ophrenia. TF-IDF was used to preprocess and vectorize text. Comparisons be-tween CatBoost and XGBoost utilized accuracy and macro precision. A second phase then applied the chosen trained model to surveys collected in order to de-termine model accuracy. CatBoost demonstrated higher macro precision, and XGBoost higher accuracy, allowing CatBoost to be chosen. 63% accuracy was achieved in the direct classification of illness, promoting high levels of success within the proposed methodology.

A Locally Deployed AI Assistant with Natural Language and Symbolic Computation Capabilities

Author Name: Zaheen Chowdhury and Muhammad Abdul Basit Ur Rahim

Abstract: Artificial Intelligence (AI) has become an increasingly large part of STEM education, being integrated through conversation assistants and tutoring systems to help with advanced problem solving. However, many AI tools rely on cloud-based large language models (LLMs) that require extensive, recurring subscription costs. This can potentially limit accessibility for students looking to use specialized AIs or even innovate and develop their own. This paper presents the development of a locally hosted AI assistant that uses an LLM deployed through Ollama with libraries for mathematics and physics problem solving. The system includes computation using SymPy and graphical representation through Matplotlib to provide step-by-step solutions, graphs, and overall interactive STEM support. Natural language interaction is also used to enhance the user experience. All of this is done without reliance on an API subscription. This paper discusses the system architecture, natural language processing, applications in educational contexts, and the advantages of local deployment - reduced cost, offline functionality, and improved accessibility. Limitations of the model and its future improvements are also discussed, including accuracy, adaptive tutoring, and an expanded range of STEM capabilities. This work demonstrates the potential locally hosted AI systems have as being cost-effective alternatives educational tools.

Analytics Framework for Course Engagement and Evaluation

Luz Diaz O'Keefe, Raquel Anden Rhoads, Dusan Ramljak

Abstract: This paper examines how limited course data can be used to study student engagement and course evaluation patterns in an online graduate course. Using aggregated Canvas engagement metrics, SEEQ/MSEEQ/SRTE course evaluation summaries, and open-ended student feedback from eight offerings of the same course, this study combines descriptive analysis of student engagement, sentiment analysis, exploratory time-series analysis, and regression. The dataset includes 93 weekly engagement records and 316 usable open-ended student responses. Results show the engagement has been generally highest early in the semester and tended to decline over time, while course evaluations remained high for most cohorts, with an overall mean score of 4.62 and term averages ranging from 4.0 to 5.0.

Lexicon-based sentiment analysis showed feedback was largely positive, with positive sentiment appearing more than 5 times as often as negative sentiment, although sentiment scores were affected by response length and the limits of lexicon-based classification. The time-series and regression are treated as exploratory because the dataset is small, aggregated, and split across discontinuous terms with varying course lengths. Overall, the study shows that limited course data can support course monitoring and future course improvement questions but should not be used to make strong predictive or causal claims.

Hyperdimensional Computing on Apple Neural Engine Chips

Author Name: Srivatsan Suresh Babu, Sophia Shahryar, and Justin Morris

Abstract: Hyperdimensional Computing (HDC) is a lightweight machine learning algorithm that is more energy efficient than deep neural networks. The algorithm leverages high-dimensional representations and matrix multiplications for classification. While there is prior work on the efficiency of HDC on specialized accelerators such as the Tensor Processing Unit (TPU), there is a lack of exploration of HDC on the Apple Neural Engine (ANE). In this paper, we evaluate the performance and energy efficiency of HDC inference across different Apple Silicon devices, such as the M3 Pro, M4, and A19 Pro chips. We express HDC inference as a two-layer neural network composed of a fixed random projection and a linear similarity layer, enabling conversion from PyTorch to CoreML for hardware-accelerated execution. Using the scikit-learn handwritten digits dataset, we benchmark inference timing and energy consumption across CPUs, GPUs, and the ANE across different batch sizes and hypervector dimensionalities. Our results demonstrate that the ANE consistently achieves better energy efficiency for large batch sizes and high-dimensional hypervectors, with up to 58% energy reduction compared to CPU execution. These findings highlight the suitability of domain-specific accelerators for HDC workloads and demonstrate the practicality of deploying HDC efficiently on modern edge devices.

Ph.D Research

Lightweight Data Lineage Tracing for Cross-Function Leak Detection in Serverless Applications

Author Name: Zuhaib Nishter and Gang Li

Abstract: Serverless platforms such as AWS Lambda compose applications from short-lived, independently deployed functions. When a single function in such a chain is compromised, sensitive data can flow silently across function boundaries to error logs, public storage buckets, or cross-tenant outputs — channels that lie outside the visibility of conventional per-function monitoring. Existing defences are either static analysis tools restricted to a single function, or runtime monitors that impose substantial instrumentation overhead, and none of them track the data itself as it moves between functions. This paper presents **LineageGuard**, a light-weight cross-function data lineage tracing mechanism that addresses this gap. LineageGuard attaches an authenticated AES-128-GCM provenance tag (sensi-tivity level, origin identifier, invocation identifier, timestamp) to each event pay-load at workflow entry, propagates it through a thin wrapper around each func-tion, and validates it at every outbound boundary using a centralized verifier im-plemented as an AWS Lambda extension. The mechanism requires no modifica-tion to existing function code and persists tags in DynamoDB. The contribution is threefold: (i) a data-bound, cryptographically authenticated provenance scheme tailored to ephemeral FaaS execution; (ii) a zero-code-modification de-ployment model using only externally documented AWS extension APIs; and (iii) a destination-policy verifier that enforces decisions on the data value rather than on indirect behavioural proxies. In a pilot evaluation across three serverless workflows — image resizing, payment processing, and log aggregation — Line-ageGuard detected all injected instances of the four leakage patterns it is designed to catch (error log exposure, cross-function token leak, unauthorized S3 write, and multi-tenant data mixing) with no false positives over 3,000 benign invoca-tions, while incurring a median latency overhead of 4.2% (range 2.8%–6.1%) and a memory overhead of less than 35 MB per function. We emphasise that these detection results are conditional on the enumerated patterns and the deployed policy specification rather than a general guarantee of perfect detection, and we report the operational and scale conditions under which they hold. The evaluation also surfaced a previously unreported error-log exfiltration path in which an au-thentication token traversed three functions before being written to CloudWatch Logs, illustrating

Epic-Organized vs. Requirement-Aligned Gherkin: An Empirical Evaluation of LLM-Based Acceptance Criteria Generation

Author Name: Shahbaz Siddeeq , Mateen Abbasi , Jussi Rasku , Zheying Zhang , François Christophe , Tommi Mikkonen , and Pekka Abrahamsson

Abstract: Automated authoring of Gherkin Behavior-Driven Development (BDD) acceptance criteria remains a manual bottleneck in requirements engineering. This study investigates whether epic-organized LLMgenerated Gherkin produces higher quality and coverage than requirementaligned generation. We compare our Timeless (an epic-organized LLM pipeline) approach against a naive large language model (LLM) baseline on four requirements documents (107 requirements) from the PURE dataset. Evaluation covers structural metrics, automated requirement coverage via TF-IDF and dense embeddings, and blind expert assessment by four researchers. In our evaluation, the JSON-constrained pipeline produced structurally valid scenarios across all generated outputs, while the zero-shot baseline achieved 99% structural validity. Semantic coverage was comparable to the baseline, with Timeless achieving 94.3% semantic Requirement Coverage Rate compared with 92.9% for the baseline.

TF-IDF produced lower coverage scores for the epic-organized output, suggesting that lexical metrics may miss coverage when scenarios paraphrase requirements at a higher level of abstraction. Expert raters prefer the epic-organized strategy on Correctness (4.61 vs 4.14), Executability (4.61 vs 4.07), and Completeness (4.31 vs 3.50). Overall, the results suggest that epic-organized generation can improve perceived Gherkin quality while maintaining comparable semantic coverage, although broader replication is needed before generalizing this finding.

An Overview of Human Aspects in the Alignment of Requirements Engineering and Testing

Author Name: Afsarah Jahin, Iflaah Salman, Rahul Mohanani, and Burak Turhan

Abstract - Background: Aligning requirements engineering (RE) and software testing (ST) involves more than just a technical perspective.
The influence of human aspects such as cognitive abilities, soft skills, and individual preferences also plays a crucial role in shaping alignment between RE and ST, underscoring the importance of investigating these aspects in this regard. Objective: This secondary study aims to provide an overview of the human aspects involved in aligning RE and ST and explores the impact of these aspects on this alignment. Method: We used a systematic mapping study (SMS) approach to explore state-ofthe-art research on the alignment of RE and ST, specifically focusing on the human aspect(s). We conducted a thematic analysis of the data to interpret the findings from the primary studies. Results: We identified five primary studies, revealing two human aspects—'communication' and 'domain knowledge'—impacting the alignment of RE and ST. Communication emerged as the most studied human aspect, exerting both positive and negative influences. Effective communication was found to be essential to achieving proper alignment between RE and ST, while any communication gap can significantly disrupt it. Conclusion: Human aspects impacting the alignment of RE and ST have received insufficient attention in the SE research community. It is crucial to conduct thorough and focused investigations in this area.

Dynamic Trustworthiness of AI for Autonomous Driving: A Design Problem Definition

Author Name: Daupadie Laknamali Ranatunga, Vishaka Basnayake, Nada Elgendy, and Tero Paivarinta

Abstract: Artificial Intelligence (AI) plays a central role in Autonomous Driving Systems (ADSs). However, current research still provides limited support for treating trust as a continuously evolving state that can be designed into the system. This paper reviews the state of the art in dynamic trust modeling for AI-assisted autonomous driving and defines a design problem for continuous trust evaluation in ADSs. A literature review was conducted systematically with three electronic databases, and the final count was seven articles. The review shows that existing work remains fragmented. The reviewed studies address dynamic trust in different ways: some model trust as a latent state in Partially Observable Markov Decision Process (POMDP) based planning, some use trust in adaptive transparency, and others represent trust as a control-related index.

However, these approaches are rarely connected to longitudinal evaluation or ADS lifecycle support. Most studies rely on simulation-based validation, focus on the Society of Automotive Engineers (SAE) Level 1-3, and operationalize trust only partially, often through POMDP/MDP inspired planning or control-oriented trust indices. Based on this analysis, we argue that the design problem has four interrelated objectives that go beyond the currently isolated solutions in the literature: (1) represent trust as a continuously updated computational state in ADS decision making, (2) infer trust online from non-intrusive behavioral and interactional signals, (3) adapt system behavior to calibrate trust without increasing driver workload, and (4) evaluate trust longitudinally across repeated human-AI interactions and system updates. The review suggests that dynamic trust in ADSs should not be treated only as a post-task evaluation issue. It also raises software engineering questions about how trust-related information can be represented, inferred, and used during system design and evaluation.

Governing LLM-Based Adaptive Components in Safety-Critical Clinical Decision Support: A Constraint-Layer Extension of MAPE-K

Author Name: Kavishwa Wendakoon1 and Nirnaya Tripathi

Abstract: Large language models (LLMs) are increasingly deployed as adaptive components in clinical decision support systems, where their generative capabilities support triage, diagnosis, and treatment planning.

The MAPE-K governance loop and its Constraint Layer mechanisms (C1-C4) provide principled runtime governance; however, these mechanisms were designed for deterministic adaptive components and carry embedded assumptions that LLMs routinely violate. We address this engineering gap by formally characterizing four LLM-specific failure modes (hallucination, sycophantic drift, distributional shift, and prompt injection) and identifying the specific C1-C4 assumption each undermines. We then derive four extensions, C1′-C4′, that retain the governance semantics of the original mechanisms while augmenting their detection logic to target these newly identified threat vectors. A proof-of-concept simulation over a nurse-led primary care triage scenario injects each failure mode under controlled conditions, replicated across two independent LLM families (GPT-4o-mini and Claude Haiku). C1′-C4′ detect three of the four injected failure categories that C1-C4 miss entirely: hallucination (TDR = 1.00, FPR = 0.14), prompt injection (TDR = 1.00), and distributional shift (TDR = 0.97). The null result for sycophantic drift (TDR = 0.00, replicated across both model families) reflects RLHF alignment resistance to short-horizon preference pressure, an informative finding with direct implications for clinical governance design. We offer C1′-C4′ as design knowledge: a lightweight constraint-layer extension that makes LLM-based adaptive behavior inspectable, controllable, and accountable in safety-critical settings, positioned as the analytical precursor to full prototype evaluation and regulatory compliance work.

HyQ-Guard – AI-Driven Runtime Anomaly Monitoring and Mitigation for Hybrid Quantum Workflows

Author Name: Ejaz Ahmed, Mikko Mäkitalo

Abstract: This paper presents HyQ-Guard, an AI-driven framework for runtime monitoring, anomaly detection, and adaptive mitigation in hybrid quantum-classical workflows. Hybrid algorithms such as the Quantum Approximate Optimization Algorithm (QAOA), Variational Quantum Eigensolver (VQE), and quantum machine learning models are central to near-term quantum computing, but their iterative execution loops remain sensitive to runtime disturbances such as shot noise, readout errors, hardware drift, and gate-level noise. Current quantum workflow platforms such as Qiskit Runtime and Amazon Braket provide execution, orchestration, and selected error-mitigation support, but they do not provide a dedicated algorithm-level runtime anomaly detection and adaptive mitigation layer for hybrid optimization workflows. HyQ-Guard is proposed as a non-invasive middleware framework that combines runtime telemetry collection with future AI-based anomaly detection and reinforcement-learning-based mitigation without modifying the underlying quantum algorithm. The objective of this initial study is not to implement the complete AI detection and mitigation pipeline, but to validate whether runtime telemetry provides measurable signals that can support such a pipeline. To establish the feasibility of the proposed monitoring layer, this paper presents an initial proof-of-concept evaluation using twelve benchmark QAOA circuits from MQT Bench, executed under controlled runtime disturbances including shot reduction, readout noise, and depolarizing gate noise. The results show that output-distribution telemetry — Total Variation Distance (TVD), Hellinger Fidelity, and output entropy — captures measurable, disturbance-specific differences between normal and disturbed executions. These findings validate the first building block of the HyQ-Guard research direction and motivate its next stage, which will incorporate complementary measures such as Jensen–Shannon Divergence alongside AI-based anomaly classification and reinforcement-learning-based mitigation for trustworthy hybrid quantum workflows.

Operationalising the EU AI Act in Software Engineering: A Systematic Mapping Study with a Focus on Healthcare

Author Name: Qaiser Khan, Abdul Raffay Saeed, and Rahul Mohanani

Abstract: The European Union Artificial Intelligence Act (EU AI Act) establishes comprehensive regulatory requirements for AI-enabled systems, with particularly stringent obligations in high-risk domains such as healthcare. Research on how to actually achieve compliance, however, is scattered across domains and varies widely in maturity. This paper reports on a systematic mapping study conducted in accordance with established software engineering guidelines. A total of 117 primary studies on EU AI Act compliance were identified and analysed, adopting a crossdomain perspective due to the limited availability of healthcare-specific contributions. The literature is systematically classified and quantitatively mapped with respect to application domains, framework types, coverage of regulatory obligations, validation strategies, and temporal trends. The results reveal a predominance of cross-domain approaches and a strong focus on governance and transparency, while comparatively limited attention is given to technical documentation, robustness, and post-market monitoring. Healthcare-specific contributions are limited, and conceptual illustration is the most common validation type, reported in over half of the studies. Based on these findings, an evidence-informed reference framework tailored to healthcare is derived as an analytical synthesis to structure compliance practices across the software lifecycle. Rigorous empirical evaluation stands out as the most pressing direction for future work.

Triangulating Prompt Engineering Effectiveness for Edge Case Detection: Integrating Expert Quality Review with Association Rule Mining in LLM-Driven Software Testing

Author Name: Usama Azhar, Saif Ur Rehman Khan, Shahid Hussain

Abstract: Large Language Models (LLMs) are being explored and proved to be useful for edge case detection in software testing. However, systematic compar-ison of different prompt engineering strategies, focusing on the quality of gen-erated edge cases, still needs to be explored. The proposed work focuses on ear-ly stage edge case detection by feeding use cases into Large Language Models (LLMs) and also investigates four prompt strategies, i.e. Simple, Role based, Contextual and Iterative that influence the generation and quality of edge cases when applied to 100 diverse set of use cases.

The generated edge cases from prompt strategies were evaluated by nine do-main experts across four quality dimensions, i.e. specificity, depth, testability and real world relevance. In parallel, Association Rule Mining was applied to uncover hidden patterns to link prompt strategies with application domains and issue types.

Results show that the iterative prompt produced edge cases with the highest complexity and richest patterns, capturing meaningful improvements across three iterations, but received the lowest expert quality scores. The simple prompt was highly effective for boundary issues (lift= 4.48). Role based and contextual prompts achieved stronger expert ratings in practical dimensions. The experimentation shows the prompt strategies significantly influence the quality of generated edge cases and demonstrates the use of prompt engineering techniques for improving edge case detection. By triangulating quantitative pat-tern analysis with qualitative expert feedback, this study aims to provide empir-ical evidence and practical guidelines for prompt selection in early stage soft-ware testing.

Reliability-Aware Retinal Encoder for Ordinal Diabetic Retinopathy Severity Modelling

Author Name: Sameera Gamage

Abstract: Retinal imaging is a useful non-invasive source of information because the retina has biological and vascular links with the brain. Most retinal AI studies focus on disease classification, but future multi-modal Neuro-AI systems also need outputs that are reliable, calibrated, and explainable. This paper presents a reliability-aware retinal neuro-oculomics encoder for diabetic retinopathy (DR) severity modelling with ordinal evaluation. DR grading is used as a supervised retinal disease-burden task with five ordered stages, where the model is trained using standard classification loss and evaluated using ordinal severity metrics. The encoder produces severity probabilities, Retinal Burden Score, uncertainty, reliability, calibration information, and explanation maps. Several backbones were evaluated using APTOS 2019 for internal testing and DDR for external validation. ConvNeXt-Tiny achieved the strongest internal result with 0.8273 accuracy, 0.8905 QWK, 0.9335 AUROC, and 0.9482 reliability. In external DDR validation, the fundus-preprocessed ConvNeXt-Tiny model achieved the best ordinal agreement with 0.6251 QWK and 0.9023 reliability. Calibration and selective prediction showed that reliability scores can help identify higher-quality predictions. Explainability analysis gave visual model-inspection support, but the maps were not treated as exact lesion segmentation. The method is not proposed as a neurological disease detector. Instead, it provides a retinal representation that can later be combined with other modalities in Neuro-AI research

A Reproducible Multi-View Mammogram Classification Baseline Using Dual-View Fusion on VinDr-Mammo

Author Name: Jitendra Sai Kota, Saddam Hossain Irfan, and Muhammad Imran

Abstract: Mammogram classification systems are often difficult to compare because studies use different label definitions, image preprocessing steps, data splits, and evaluation rules. This lack of agreement makes it hard to judge whether a new model is truly better or only benefits from a different experimental setup. This paper presents a clear and reproducible baseline for breast-level mammogram classification on the VinDr-Mammo dataset. The study focuses on a practical clinical question: whether a breast should be treated as routine or should receive further attention. We map BI-RADS 1 and 2 to the routine class and BI-RADS 3, 4, and 5 to the attention class. We compare single-view classification with a dual-view model that uses both craniocaudal (CC) and mediolateral oblique (MLO) views of the same breast. The model uses a shared EfficientNet-B5 backbone and averages the learned CC and MLO embeddings before classification. We also test three contrastive learning objectives. Across five study-level cross-validation folds, dual-view fusion improves mean AUC from 0.784 to 0.795 and improves mean F1-score from 0.432 to 0.456. Contrastive losses give small AUC changes but no consistent practical gain. The project repository is available at https://github.com/iCARE-Lab/vindr-mammo-dual-view-classification.

Supporting Reflection on Sustainable Software Design Decisions in Software Engineering Education: An Evaluation of a Socratic Chatbot

Author Name: Kseniia Suomalainen, Elena Cavallin, Viktoriia Olshanskaia1, Maria Luce Lupetti, and Jari Porras

Abstract: Early software design decisions can influence the sustainability of software systems. However, software engineers often make such decisions under uncertainty, and their reasoning may be shaped by cognitive biases that limit the consideration of long-term environmental impacts. Despite growing interest in sustainable software engineering, tools that support reflection on sustainability-related cognitive biases remain limited. This paper evaluates of BIAS 3v, a Socratic conversational agent designed to support reflective dialog about software design decisions. The chatbot guides software engineering students through adaptive questioning aimed at uncovering cognitive biases and encouraging deeper reflection on sustainability trade-offs in their projects. We conducted a preliminary mixed-methods study with 12 software engineering students from Finland and the Netherlands, combining a pre-questionnaire, interaction with the chatbot, and a post-questionnaire. Preliminary results suggest that students who used this tool appeared more capable of recognizing cognitive biases conceptually, while still tending to underestimate their own susceptibility to biased reasoning. Additionally, participants showed better sustainability awareness and stronger ownership of sustainability concerns. This work contributes an exploratory evaluation of an adapted tool and provide insights into AI-driven metacognitive scaffolding to support bias-aware reflection for software engineering education.

Analytics Framework for Course Engagement and Evaluation

Author Name: Luz Diaz O'Keefe, Raquel Anden Rhoads, Dusan Ramljak

Abstract: This paper examines how limited course data can be used to study student engagement and course evaluation patterns in an online graduate course. Using aggregated Canvas engagement metrics, SEEQ/MSEEQ/SRTE course evaluation summaries, and open-ended student feedback from eight offerings of the same course, this study combines descriptive analysis of student engagement, sentiment analysis, exploratory time-series analysis, and regression. The dataset includes 93 weekly engagement records and 316 usable open-ended student responses. Results show the engagement has been generally highest early in the semester and tended to decline over time, while course evaluations remained high for most cohorts, with an overall mean score of 4.62 and term averages ranging from 4.0 to 5.0. Lexicon-based sentiment analysis showed feedback was largely positive, with positive sentiment appearing more than 5 times as often as negative sentiment, although sentiment scores were affected by response length and the limits of lexicon-based classification. The time-series and regression are treated as exploratory because the dataset is small, aggregated, and split across discontinuous terms with varying course lengths. Overall, the study shows that limited course data can support course monitoring and future course improvement questions but should not be used to make strong predictive or causal claims.

Autonomous Multi-Agent Governance: AI Sentinel Framework for Mitigating Psychological Restraints in AI-Driven Fall Management

Author Name: Tushar, Anushka Thagle1, Supanut Chindawan, Sanjna Sai Chintakunta, Muhammad Faizan Raza, Albert Elcock, Rich Moore, Praneeth Sunkavalli, Chandan Shivalingaiah, Gaurav Prabhu, Lukas Sparer, Alexander Neulinger, Sanjeev Sondur, Arturo Casasa, Maryam Roshanaei, and Dušan Ramljak

Abstract: In the rapidly evolving landscape of geriatric healthcare and its governance, fall-management industry members like our industry partner MyRaina, Inc. are seeking proactive, technology-driven solutions that safeguard vulnerable residents while preserving their autonomy and ensuring regulatory compliance. To reconcile clinical, ethical, and legal demands in a shifting regulatory environment, we leverage AI Sentinel, a multi-agent framework under development, as an autonomous trust layer to enforce ethical controls and regulatory compliance in fall management. MyRaina's fall-monitoring system is our motivating case study/target deployment environment, while deterministic rule-based prototype is a starting point towards a fully autonomous multi-agent governance system. By using a meta-agent to coordinate specialized sub-agents responsible for auditing all components of the surveillance system, we identify behavioral indicators for restrictions. The system's output, comprising immutable audit logs, indicates the guardrails, guidelines, and best practices required for fall-management systems to detect potential non-compliance related to restraints. We have completed the first of a four-milestone roadmap, spanning: 1) generating data leveraging real production schemas, 2) retrospective analysis under a data-use agreement, 3) shadow deployment, 4) IRB-supervised clinical validation. Our results show that this multi-agent approach can transform passive risk technology into a transparent, human-centered system characterized by explainable insights, privacy-by-architecture, and a local-first design. In an industry continually balancing regulatory compliance with obligations to protect and support residents, AI Sentinel provides early detection of Tushar et al.