2026 IEEE Cloud Summit

Washington, District of Columbia, USA
25-26 June 2026

IEEE Catalog Number: CFP26U01-POD
ISBN: 979-8-3315-8196-1

IEEE
Advancing Technology for Humanity

Copyright © 2026, IEEE

All Rights Reserved

Copyright and Reprint Permissions:

Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

*** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.

IEEE Catalog Number:	CFP26U01-POD
ISBN (Print-On-Demand):	979-8-3315-8196-1
ISBN (Online):	979-8-3315-8195-4

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA

Phone:	(845) 758-0400
Fax:	(845) 758-2633
E-mail:	curran@proceedings.com
Web:	www.proceedings.com



TABLE OF CONTENTS

Frontier-to-Runbook: Compiling Censor-aware Guardrails for Multi-tenant Telemetry Storage 1
Fuming Fu

Designing Failure-Resilient AI Pipelines Under Continuous Infrastructure Churn in Cloud Environments 7
Varun Raj

Dynamic Prompt Routing via Model Context Protocol: Evaluating LLM-Based Template Selection for Content Safety 11
Gowtham Shankar, Changsha Ma, Rakesh Shirke

Cloud-Native Reference Architecture for Agentic Commerce: AI Shopping Agents in Payment-Enabled Platforms 17
Rakesh Shirke, Gowtham Shankar

Understanding When Declarative Feature Definitions Are Beneficial: Observations from Feature Materialization 23
Bhala Ranganathan, Mickey Zhang, Feng Pan

What Academic AIOps Research Gets Wrong: A Practitioner's Perspective from 12 Years of Cloud Operations 28
Mateen Ali Anjum

ANQOR: Anchor-based Quantum Multi-class Image Classification 34
A K M Muhitul Islam, Yafan Huang, Yan Yan, Guanpeng Li

Towards Building Reliable Machine Learning Surrogates for Scientific Applications 41
Bohan Zhang, Wenqian Dong, Guanpeng Li

COSMOS-R²: Artificial Intelligence–Driven Risk-Bounded Carbon-Constrained Autonomous Cloud Scaling via CMDP Control 49
Sadhasivam Mohanadas, Leela Prasad Gorrepati

Quantization Undoes Alignment: Bias Emergence in Compressed LLMs Across Models and Precision Levels 57
Plawan Kumar Rath, Rahul Maliakkal

Building An Elastic, Resilient, Scalable Cloud-Native Digital Twin by Leveraging Stream Adaptation Agents for Smart Distributed Manufacturing Systems 64
Prachi Kedar, Vamsi Mulpuri

Adversarial Resilience Engineering for Digital Finance Microservices: Stability Controls and Amplification-Aware Incident Replays 70
Sibasis Padhi

A Single-Node Parallel Framework for Bulk Binary File Extraction from Salesforce CRM: Design and Performance Evaluation 74
Bhimeswara Vamsi Punnam

MCP-Edge: Extending the Model Context Protocol to Constrained IoT Devices 80
Muntaser Syed, Sheikh Abujar, Sharun Akter

Evaluating Language Models on Infrastructure-as-Code Development 86
Natalia Tarasova, Enrique Balp-Straffon, Aleksei Iancheruk, Yevhenii Sielskyi, Nikita Kozodoi Liam Byrne Jack Butler, Dayuan Jiang, Marcin Czelej, Andrew Ang, Yash Shah, Roi Blanco, Sergei Ivanov

EWOK: A Flexible Agent-Based Disaster Recovery Orchestration System for Stateful Cloud Services 93
Sahil Sabharwal, Akshay Pratinav, Chaitanya Bhatt

Load-Aware Offline Scheduling for Heterogeneous Edge Deep Learning Inference 98
Elvis G. Fefey, Tanzima Z. Islam

Exploring Information-Freshness v. Task-Latency Tradeoffs for Schedulers in Chiplet-based Hardware 104
Ege Dogan, Ali Suvizi, Egemen Erbayat, Suresh Subramaniam, Guru Venkataramani

Centralized MLOps Platform for Software Defined Vehicles 110
Pavan Vemuri

Cold Start Latency of Quantized Small Language Models on Serverless CPU Infrastructure 116
Ming Hu, Khushan Adatiya

Semantic Control Plane for Enterprise Generative AI in Cloud Environments: Architecture, Governance, and Production-Domain Evaluation 122
Leonardo Gomez, Diego Ortiz, Jose Romero, Nishchai JM, Rui Qing Zhang

Beyond NL2SQL: Integrating Model Context Protocol with Oracle Select AI for Agentic Business Intelligence on Autonomous Database 128
Chaitanya Kulkarni, Rohit Nimmala, Chandrashekhar Medicherla, Siva Prasad Nandi, Milan Parikh Ram Hibane

CAPS: A Transformer-Based Financial Sentiment Forecasting Framework for Multi-Asset Equity Return Prediction 134
Darryl Carpenter, Boxiang Dong, Bharath K. Samanthula, Weifeng Sun

Cache-Aware Auto-Suspend Control for Elastic Cloud Data Warehouses under Quantum Billing Constraints 140
Avik Datta, Sai Pranav Vuppala

Cloud-Native Graph-Aware Federated Learning for Diabetic Retinopathy Diagnosis: Architecture, Orchestration and Graph Aggregation 148
Jesu Marcus Immanuvel Arockiasamy

Mitigating Confused Deputy Attacks in AWS Cross-Account IAM Roles Through ExternalId Enforcement 154
Ricky Gole, Vahid Heydari

Production-Scale Cloud Deployments and Operations on AWS Using EKS 161
Mukesh Aurangabadkar, Harish Apuri, Shikher Goel, Uudit Misra

Health Data Analysis: Definition, Evolution, and Future Prospects 168
Bingwei Liu, Marinka Natale

DRFU: Federated Unlearning with Reverse KL-Divergence-based Knowledge Distillation 174
Thanh Nam Nguyen, Hai Anh Tran, Huy Hieu Nguyen, Nam Thang Hoang, Yixin Zhang, Mahima Konda Truong X. Tran

Dynamic Task Allocation and Path Planning for Multi-UAV Disaster Search Using Multi-Agent Reinforcement Learning 180
Malek Chabbouh, Amr Mohamed

Sensor-Reactive UAV Search and Rescue via Deep Reinforcement Learning 186
FatemaElzahraa Elrotel, Noor Khial, Mhd Saria Allahham, Loay Ismail, Amr Mohamed

Walk Carefully: Hyperparameter Sensitivity of Node2Vec for Security Event Graph Clustering 192
Moazzam Khan, Asad Waqar Malik, Mohammad Omer

Benchmarking GPU Performance Characteristics of Graph Foundation Models 198
Chaoqun Li, Runbang Hu, Joseph Caldwell, Xiaojiang Du, Yuede Ji

Beyond CVSS: Context-Aware Exploitability Assessment of Vulnerabilities in HPC Containers 204
Emmanuel Emakhue Oseghale, Mohammad Shahedur Rahman, Ashwani Manoj Kumar Rai, Xing Gao Yuede Ji

SHIELD: Scalable HTTP Integrity Enforcement for Large Deployments 210
Osamah Alzacko, Shamil Al-Ameen, Latifur Khan

Nonlinear Feedback Control for Cyber Resilience: The Security Feedback Control Architecture (SFCA) 216
Atharva Khambete, Avanish Kulkarni, Bharati P. Vasgi, Punam Chavan

The Context-Aware Data Fabric (CADF): Dynamic Context Pruning for Cost-Optimized Agentic AI Workflows 221
Gautami Nadkarni

Distributed PostgreSQL as an IDP Primitive: Multi-Region Topology and Tail-Latency Remediation 225
Gaurav Saxena, Agustin Reyna

Autonomous Agent-Based Observability Framework for Cloud-Native Systems - SENTINEL 229
Santosh Srinivasan, Giridhar Shyam Sankararaman, Aruun K. Kumar

HDPRO: Preventing Orphaned Bursts with Deterministic Sub-500ms Recovery in Kubernetes 235
Amandeep Singh, Aamani Nemtur

GitOps-Driven Adaptive Closed-Loop Multi-Cloud Governance Framework for Drift Correction, Policy Enforcement, and Self-Healing Operations 240
Arun Pandiyan Perumal

Exploring Computational Energy Loss in Connected Vehicles 244
Asad Waqar Malik, Samee U. Khan

Evaluating the Impact of a Semantic Layer on LLM-Based Text-to-SQL Accuracy for Enterprise Data Querying: An Empirical Study 248
Diego Ortiz, Leonardo Gomez, Jose Romero, Nishchai JM, Rui Qing Zhang

AI-Driven Data Quality Monitoring for Enterprise Cloud Data Platforms 252
Saqib Khan

LightKubeGuard: A Lightweight Machine Learning Framework for Detecting Resource Hotspots in Kubernetes-Based Cloud Systems 258
Digdarshan Subedi, Karma Gurung, Nikesh Tiwari

AI in Healthcare Operations: Definition, Evolution, and Future Prospects 264
Christine Miyachi

Safe and Efficient Resource Allocation in Multi-AI-Agent LLM Systems 268
Dhwanil M. Mori, Neil F. Johnson

Serverless, Unmasked: A Decade of Footprint, Cold-Starts, and Control-Plane Insights 271
Sabyasachi Gupta, Charles Dailey, Paul Gratz, John Lusher