GENERATIVE AI WORLD SUMMIT with MLOPS WORLD CONFERENCE

5th Annual

MLOPS WORLD CONFERENCE & EXPO

www.mlopsworld.com

GENERATIVE AI WORLD SUMMIT

www.generative-ai-summit.com

Joint-Event: November 7-8th, Austin TX, Renaissance Austin Hotel

YouTube

LinkedIn

5th Annual MLOps / GenAI World Conference & Expo

2024 Committee Topics

Thank you for taking an interest in our 5th annual MLOps/GenAI World Conference and Exhibition. We look forward to growing our ML communities and continuing our efforts to build and strengthen our community together.

Our Mission **Our annual gathering will take place this year, in Austin Texas, November 7-8 2024: https://mlopsworld.com/ alongside https://generative-ai-summit.com/**

MLOps World is an international community group of practitioners trying to better understand the science of deploying ML models into live production environments. Our goal is to further the understanding of best practices, methodologies, principles around ML/AI/GenAI experimentation and deployment. Our community consists of over 15,000 members exploring best practices, methods, and principles around ML/AI/GenAI in production environments.

Meet Event Committee & Team

See Who Attends

The conference will be held in person, Austin Texas amongst a beautiful setting 15 mins north of Downtown Austin. Situated within the large hotel facility and next to many outdoor attractions, the event will include various opportunities to meet and share experiences through planned outdoors outside of the main conference sessions and workshop hours.

Speaker Submission page https://sessionize.com/mlops-generative-ai-world-2024/

- **Formats:**
  - **15 min - 45 min presentations**
  - **1.5 hr - 3 hr workshops.**
  - **45 min-1.5 hr discussion room**
- **IMPORTANT DATES**
  - **Deadline for submission of papers: July 31st 2024**
  - **Conference Dates: November6 (virtual day) November 7-8th (in-person) 2024**
- **CONTACTS:**
  - **Executive Organizer: David Scharbach: info@mlopsworld.com**
  - **Industry booth participants: Faraz@mlopsworld.com**
  - **General Inquiries including Visa related inquiries: info@mlopsworld.com**

2024 Committee Topics

(Not limited to)

Pain points:

- Infrastructure and Deployment
    - Developments in Agentic Ops
    - Using Internal Legacy Infrastructure to Train and Deploy ML Models, as an Alternative to Cloud
    - Deployment and Monitoring
    - Managing GPU Clusters
    - Deploying Multimodal Models in Production
    - Overcoming Challenges to Perform Production-Grade Deployment of Models while Scaling Across Cloud Environments
    - Deploying Edge Computing, Heterogeneous ML Pipeline Orchestration, and Digital Twin Technology
    - Inference and Training on the Edge
- Scalability and Performance Optimization
    - Scaling Data and Model Pipelines
    - Scaling Workflows, Tracking Lineage, and Ensuring a Smooth Dev Experience
    - Model Training and Serving at Scale with GPU
    - Distributed Training of Large Models
    - Scaling Up LLMs
    - Reducing Inference Costs
- MLOps and Lifecycle Management
    - Moving from Experimentation and Proof-of-Concept to Production
    - Making Complex MLOps Setups Easy for Engineers with Little ML Experience
    - Moving Through MLOps Level 0 to Level 1 and Level 2
    - Addressing Challenges Associated with E2E Lifecycle of CV/ML Models Including Dataset Creation and Model Validation, Regression Testing, Diagnostics and Debugging
- Cost Optimization
    - Reducing Cost of Running in Production
    - Optimization of Cost Associated with Infrastructure of Model Training and Inference
- Data Management and Quality
    - Data Annotation
    - Accessing and Managing Large Scale of Datasets in Generative AI Use Cases
    - Limited Amount of Labeled Data and Privacy-Preserving Training

    - Improving ML Techniques Through Data Augmentation
- Model Selection, Training, and Fine-tuning
    - Continual Learning for Fine-Tuned Model to Adapt to the Growing (and Sometimes Drifting) Data
    - Improving Transformer Architectures and Comparing Performance with Current SOTA Results
    - Executing LLM Based Approaches Through Prompt Engineering or Fine-Tuning
    - Several Situations Where LLMs are Replacing/Augmenting Older ML Techniques: How to Compare Different Approaches
- Security, Privacy, and Compliance
    - Meeting Security and Compliance Requirements Across the Entire Lifecycle of ML Projects and Tooling
    - Enabling Security Patching for Open Source MLOps Tooling
    - Overcoming Barriers to Generative AI Solutions Due to Data Privacy Related Issues
- Model Evaluation and Governance
    - Models in Production, Evaluation
    - Better Execution Model Validation, Governance, and Training
    - Discovering the Decisive Features that Led to a Diagnostic Result
- Integration and Interoperability
    - Bridging ML Infrastructure/Platform Work to Company Product Impact
    - Increasing Relevance of Search Across Different Platforms
    - Better Prompt and LLM Support by Big Cloud System
    - Anticipating Challenges with Current Single-Model Infrastructure While Expanding to Support Multiple Languages
    - Getting LLMs to Work with API Specifications
- Business Alignment and Industry Adoption
    - Optimization of Digital Ads Campaign Strategies
    - Reusability and Scalability of Previously Proven ML Solutions for Similar but New Problems
    - Preparing Traditional Industry to Embrace the Change of AI
    - Stakeholder Business Alignment for AI Applications
    - Stakeholder Trust in Model Results
    - Finding an Easy Way to Connect ML Infrastructure/Platform Work to Company Product Impact

Listed Areas of Committee Interest:

Generative AI & Large Language Models (LLMs):

- Commercializing Generative AI for ML Professionals
- Implementing Generative AI Use Cases in Production
- Effectively Fine-Tuning Large Language Models
- Best Practices for Fine-Tuning Larger Models for Specific Applications
- Exploring Creative Applications of Large Language Models
- Deploying and Managing LLMs/GenAI Systems in Production Environments
- Addressing MLOps/LLMOps Challenges in Generative AI Use Cases
- Designing Agentic AI Applications and Frameworks for Enterprise

MLOps, LLMOps, Agentic Ops Infrastructure:

- Developments in Agentic Ops
- Implementing End-to-End AI and ML Deployment Processes
- End User Case Studies Around Developing and Implementing MLOps Lifecycles
- Implementing a Modern MLOps Stack
- Automating MLOps Processes with LLMs
- Enhancing Observability in MLOps and AI Systems
- Monitoring and Observing Multimodal Generative Models
- Ensuring Continuous Delivery with MLOps Workflows
- Comparing MLOps and Traditional Ops: Current Challenges and Solutions
- Adopting MLOps Best Practices in AWS, GCP, and Azure
- Exploring the Intersection of MLOps and Data Engineering
- Utilizing Pyspark/Scala for Distributed ML Compute
- Implementing Real-Time ML and Distributed Training/Inference
- Scaling Inference and Deploying Heterogeneous Systems
- Conducting Scoring and Inference in Streaming Mode
- Using Vector Databases for Real-Time ML Inference
- Scaling Vector Databases with Quantization Techniques
- Achieving Real-Time Inference with LLMs
- Optimizing Models and Conducting Online Testing
- Optimizing GPU Usage and Costs for ML Workloads
- Accelerating ML with Hardware-Agnostic Techniques
- Implementing Cybersecurity Best Practices for ML and GenAI in Production
- Ensuring Privacy and Security with Federated Learning
- Re-Training and Monitoring Models Effectively
- Managing the Lifecycle of LLMs and Multimodal Models
- Identifying and Implementing Effective RAG Architectures
- Applying Meta Learning and Graph Neural Networks in Production

- Following Best Practices for Model Testing and Validation
- Adapting to Events that Challenge ML Model Accuracy
- Building MultiModal MLOps/LLMOps Pipelines and Infrastructure
-

Practical Applications & Industry-Specific Insights:

- Simplifying and Automating MLOps with LLMs
- Implementing Multimodal Use Cases in Production
- Scaling ML Model Training (Including with Spark and Ray)
- Efficiently Serving ML Models at Scale
- Optimizing and Sustaining AI Model Training and Inference
- Ensuring Performance Monitoring Post-Deployment
- Incorporating Privacy by Design in ML Pipelines
- Driving Innovations in Generative AI
- Leveraging Generative LLMs for Code Completion (e.g., GitHub Copilot)
- Utilizing General-purpose Prompt-answer Systems like ChatGPT
- Exploring Quantum Computing's Impact on ML
- Bringing Open Source and Academic Advancements to Enterprise
- Sustaining New AI Products with Auxiliary AI Models (AI for AI)

Technical & Operational Challenges:

- Implementing Federated Learning for IoT ML Projects
- Engineering Perspectives on Infra Costs for Projects like ChatGP
- Best Practices for Application Models with Dynamic Infrastructure
- Supporting Women and Underrepresented Groups and Promoting Diversity in ML Teams
- Ensuring AI Safety in Production Environments
- Making Machine Learning Financially Viable
- Leadership Challenges in the ML World
- Bridging Advancements in Academia and Open Source into Enterprise Production
- AI/GenAI Strategy and Product Development
- Best Practices for Designing Applications Using Different Models
- Infrastructure for Loading Models On-demand Based on Specific Use Case Requirements
- Automated LLM Model Evaluation at Scale Use Cases and Best Practices
- Running Models in Hybrid Cloud Environments
- Strategies to Increase Efficiency and Reduce Carbon Footprint of Large Models
- Increasing Team Efficiencies and Effectiveness Using LLMs Internally
- Managing Employee Adoption of AI Tools

Democratizing AI & Ethical Considerations:

- Democratizing AI in Enterprise Environments Beyond ML Practitioners
- Ensuring Transparency and Data Privacy in AI Solutions
- Establishing Ethical Standards and Accountability in ML Model Evaluation
- Implementing Ethical AI Practices with Practical Tools
- Ensuring the Responsible Use of Generative AI
- Classic ML vs. Generative AI Use Cases Beyond RAG

Not sure what to submit or have any questions ? Email us at info@mlopsworld.com

- Link to submit your talk
- 2023 Event Video

- Our Socials