

AI INFRA SUMMIT

ENTERPRISE AI INSIGHTS

INCLUDING



ALAN ROSA, CISO & SVO, TECHNOLOGY, CVS HEALTH

KEYNOTE

LESSONS LEARNED FROM ROLLING OUT ENTERPRISE-WIDE AI

DAVID GLICK, SVP, ENTERPRISE BUSINESS SERVICES, WALMART AND ATIF RAFIQ, FOUNDER, RITUAL

01 Rapid Iteration

AI enables significantly faster iteration in the software development lifecycle, allowing prototypes to be built and refined in hours or days instead of months.

02 Curiosity as a Skill

Success with AI is driven by curiosity, persistence, and resilience, rather than just traditional technical skills, as roles become more fluid.

03 Streamlining Governance

Walmart adapted existing robust governance processes for AI, and actively worked to remove bottlenecks, such as reducing security approval times from months to zero days.

04 Building Trust through Delivery

Rapid delivery of AI solutions helps build trust between business, product, and tech teams, encouraging faster adoption and collaboration.

05 Future of Software Development

Code writing will become a vanishingly small part of the SDLC, with AI automating upstream (PRDs) and downstream (testing, security reviews) processes, limited only by creativity.



ATIF RAFIQ, FOUNDER, RITUAL



DAVID GLICK, SVP, ENTERPRISE BUSINESS SERVICES, WALMART



PANEL

ADVANCING AGENTIC PERFORMANCE WITH ENGINEERING INNOVATIONS

01 Agentic Projects

Crusoe is shifting to inference, providing tools and abstracting infrastructure for agentic workflows, while Cloud Code helps deploy agents for technical users.

02 Capital One's Multi-Agent Approach

Milind details Capital One's use of multi-agent systems for enterprise applications, starting with conversational AI, emphasizing the need for accuracy, regulation, and cost/latency optimization.

03 Hardware Determinism & Precision

Jill highlights the importance of deterministic hardware and high-precision math (e.g., 16-bit) for agent trust, as it impacts reproducibility and decision quality, especially when evaluating complex agentic systems.

04 Infrastructure Stack Changes

The panel discusses the need for frameworks to incorporate locality awareness (e.g., within MCP) to reduce latency hops between models, especially across diverse chipsets and data centers.

05 Vertical Integration & Optimization

Milind emphasizes deep customization of LLMs for each agent and vertical integration to optimize the inference stack, reduce latency, and distill models for cost efficiency, citing 15 LLM invocations per user interaction.

06 Cost & Environmental Impact

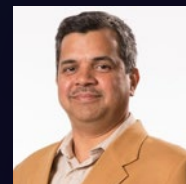
Crusoe engineers for sustainability by using renewable energy sources for data centers, while others focus on optimizing prompts and system efficiency, believing that technical efficiency naturally leads to lower costs and environmental impact.



Gilles Backhus

Founder & VP, Product
Tensordyne

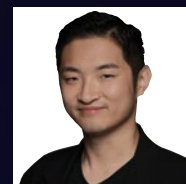
TENSORDYNE



Milind Naphade

SVP, AI Foundations
Capital One

Capital One



Dickson Tsai

MTS, Claude Code
Anthropic

ANTHROPIC



Kyle Sosnowski

VP, Software Engineering
Crusoe

Crusoe



Vamsi Patti

Investor
Hitachi Ventures

HITACHI
Ventures

SELF-HOSTED VS. PROPRIETARY... WHAT ARE THE TRADE-OFFS?

ARUN NANDI, CHIEF DATA & AI OFFICER, **CARRIER**

- 01 Proprietary Model Growth**
Proprietary models have seen rapid consumer adoption (e.g., ChatGPT, DeepSeek) and significant enterprise penetration, with OpenAI claiming 3 million enterprise seats and Anthropic 300,000 customers.
- 02 Proprietary Model Advantages**
Proprietary platforms offer comprehensive out-of-the-box features like governance (RBAC, SSO, compliance), security (IP, data privacy), observability, and performance optimization tools (prompt caching, query routing).
- 03 Open Source Performance & Adoption**
Open-source models are catching up in performance, democratizing AI capabilities, and currently account for approximately 20% of AI workloads.
- 04 Decreasing Token Costs**
There has been a significant decrease in token costs (around 97% in the past 1.5 years) for flagship models, with a trend towards near-zero inference rates.
- 05 Open Source Deployment Options**
Open-source models can be self-hosted or accessed via APIs through co-hosted services, offering cost benefits compared to proprietary models for similar performance.
- 06 Hybrid Future & Recommendations**
The future of AI model deployment is hybrid, leveraging both proprietary (for high-impact, customer-facing use cases) and open-source models (for innovation, IP, data sovereignty, edge). Recommendations include trying self-hosting open-source models, establishing governance, evaluating price/quality, and planning for a hybrid strategy.



Carrier

KEYNOTE PANEL

MANAGING SUCCESSFUL FIRMWIDE AI INFRASTRUCTURE STRATEGY & DEPLOYMENT

01

AI Application in Pharma

Pharma has a long history of using AI/ML for drug discovery, now expanding with GenAI into back-office operations and more sophisticated modeling, requiring process redefinition and investment in infrastructure and talent.

02

AI Application in Healthcare

CVS Health leverages AI across its diverse segments for back-office efficiency, optimizing claims processing, accelerating drug supply chain management, and deploying intelligent IVR solutions for customer service.

03

Managing Expectations and Risk

There's pressure for ROI from stakeholders, requiring education about initial high failure rates. Highly regulated industries like pharma and healthcare face immense responsibility regarding data security (PHI/PII), patient safety, and avoiding bias.

04

Security and Privacy Concerns

Key security considerations include securing data sources for model training, authorizing internal and third-party models, and adhering to principles of trust and safety like transparency, explainability, and bias mitigation.

05

Talent Evolution

The industry is seeing a shift towards more formalized training and a higher barrier to entry for AI roles, attracting talent from high-tech backgrounds. Companies are focusing on attracting new talent, retraining existing workforce, and exploring non-traditional talent sources like apprenticeship programs.

06

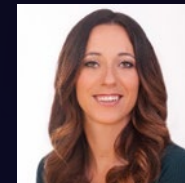
Infrastructure Challenges

Increased demand for compute, especially in the cloud, is a continuous trend. The need for higher density in data centers, managing power/cooling (e.g., water cooling), and addressing network bottlenecks (e.g., 10 gig vs. 800 gig interfaces) are significant infrastructure challenges.



Alan Rosa

CISO & SVO, Technology
CVS Health



Berta Rodriguez-Hervas

Chief AI and Analytics Officer
Pfizer



Robert Brooks

Founding Team & VP Revenue
Lambda



EXCLUSIVE INTERVIEW WITH...

DAVID GLICK, SVP, ENTERPRISE BUSINESS SERVICES, **WALMART**

