

AI-Infused Software is Eating IoT's Edge

James Kobielus

Lead Analyst, SiliconANGLE Wikibon

 PentahoWorld

Agenda

Edge-based AI is the most disruptive trend in modern application development

- AI is Data-First Application Development
- The IoT Edge Is Where The Most Disruptive AI Will Live
- Developers Should Design Decoupled AI for Edge Deployment
- AI Algorithms Must Conform to Edge Resource Constraints
- DevOps Practices Are Key to Edge AI Governance
- Summary
- Next Steps

AI is Data-First Application Development

- AI consists of machine learning, deep learning, and other data-driven algorithms
- AI augments users' organic powers of cognition, reasoning, natural language processing, predictive analysis, and pattern recognition.
- AI-driven digital assistants drive smarter decisions in commerce, mobility, messaging, social, and other applications.
- Well-engineered AI accurately predicts desired outcomes and understands user intentions
- Self-learning AI adaptively refines algorithms from fresh data, user interactions, and changing environmental, social, and other contexts



The IoT Edge is Where the Most Disruptive AI Will Live

- AI is eating IoT's edge through embedding as a core capability of all endpoint nodes and applications.
- In the IoT, embedded AI processes the rich streams of real-time machine data being captured by edge devices
 - E.g., smart thermostats, commercial drones, self-driving vehicles, and industrial sensors.
- Embedded AI imbues edge devices and apps with their core smarts
 - E.g., situational awareness, video recognition, motion detection, natural-language processing



Developers Should Design Decoupled AI for Edge Deployment

- Monolithic AI development is out of sync with the radically distributed IoT edge fabric.
- Developers should decouple AI functions as modular microservices that can be deployed over federated cloud-computing environments to edge devices
- Implement real-time AI functions primarily on edge devices and gateways, thereby reducing or eliminating the need to round-trip to the cloud
- Containerize AI functions across edge, gateway, and cloud nodes, enabling orchestrated execution of complex application across IoT cloud fabrics



AI Algorithms Should Conform to Edge Resource Constraints

- Handle in-memory, real-time, and low-latency workloads involving locally-acquired sensor data
- Execute compute-intensive hierarchical tasks (e.g., image, video, and audio recognition)
- Optimized for ASICs and other custom high-performance chips
- Incorporate simpler feature spaces and fewer independent variables
- Operate in intermittently connected, low-bandwidth, autonomous-decisioning scenarios



DevOps Practices Are Key to Edge AI Governance

- Manage all edge-AI algorithms, models, code, and other pipeline artifacts within a centralized source repository
- Implement a IoT-optimized data lake for management of edge-AI data for modeling, visualization, training, refinement, auditing, compliance, and governance
- Deploy a unified cloud platform for team-based collaboration in modeling, training, deployment, evaluation, and other edge-AI development tasks
- Enforce consistent policies for sharing, reuse, permissioning, check in/check-out, versioning, training, deployment, monitoring, and other governance requirements for all edge-AI projects



Summary

What we covered today:

- Developers should be prepared to embed AI software into IoT endpoints.
- Doing so will enable these edge nodes to make decisions and take actions autonomously based on algorithmic detection of patterns in locally acquired sensor data.
- Decouple and deploy AI functions as modular IoT microservices that
 - fit the resource constraints of edge devices,
 - can be deployed over federated cloud-computing environments to edge devices, and
 - can be governed centrally, automatically, and remotely over their lifecycles
- Don't forget to engineer downstream edge-AI application/algorithm governance for extreme scalability

Next Steps

Want to learn more?

- Industry Initiatives Pushing AI-Infused Software to the Federated Edge: <https://wikibon.com/industry-initiatives-pushing-ai-infused-software-federated-edge/>
- Building AI Microservices for Cloud-Native Deployments: <https://wikibon.com/building-ai-microservices-for-cloud-native-deployments/>
- Agile Development in Team Data Science: <https://wikibon.com/agile-development-in-team-data-science/>
- Optimizing Your Application Architecture At The Federated Edge: <https://wikibon.com/optimizing-your-application-architecture-at-the-federated-edge/>



 PentahoWorld