Yan-Cheng (Bill) Hsu

| profile.bill-hsu.com | bill.ych.jobs@gmail.com | www.linkedin.com/in/yan-cheng-hsu | (+1)858-319-5739 |

PROFESSIONAL EXPERIENCE

Site Reliability Engineer II | **Alibaba Cloud** – AnalyticDB Org – AI Training Platform Resource Management | Sunnyvale, CA, U.S. | Jul. 2025 – Present

Project Nexus: Cross-Cluster AI Training Infrastructure for Unitree G1-D(unitree.com/G1-D)

Unified Resource Orchestration (System Design): Designed the **Dual-Layer Virtual Kubelet** architecture to centralize **10,00+ heterogeneous GPUs**, enabling a migration from Serverless to Reserved instances that achieved **~40% cost reduction** and scaled Unitree G1 Robot training capacity by **25x**.

- **Federated Identity Mesh:** Designed and implemented a novel credential injection mechanism by **mocking Service Accounts** via externally-mounted Secrets. Enabled secure **Cross-Cluster AuthN/AuthZ**, allowing remote Pods to authenticate with the Master API Server as if local, solving "split-brain" identity challenges using Custom Controllers with 9-hour token rotation.
- **Hybrid Network Fabric & Service Discovery:** Automated VPC networking via **Terraform** (Security Groups, CLB ACLs) to overcome CoreDNS isolation. Re-architected Ray's service discovery by mapping internal endpoints to **CLB External IPs**, ensuring low-latency Head-Worker communication across network boundaries.

AIOps Hybrid Runtime Observability & Self-Healing (Infrastructure Automation): Architecting a telemetry pipeline and closed-loop controller to automate fault recovery for isolated GPU sandboxes, targeting a significant reduction in Mean Time to Resolution (MTTR) for training interruptions.

- **Secure Enclave Telemetry Pipeline:** Architecting the data exfiltration strategy to integrate SysOM agents via Helm within a restricted mixed-runtime environment. **Configuring** critical network paths (Security Groups, Firewalls) to tunnel **DCGM metrics** and kernel traces from isolated sandboxes to SLS/OSS sinks, ensuring zero data loss.
- **Autonomous Remediation & Stability Engine:** Developing a custom **Kubernetes Controller** to consume SysOM diagnostic APIs. **Programming** the logic to ingest real-time anomaly signals and trigger autonomous self-healing workflows (cordon/drain) for unhealthy GPU nodes, creating a resilient closed-loop system.

Software Development Engineer | **Amazon** Artificial General Intelligence Org – High Performance Computing | Seattle, WA, U.S. | Oct. 2023 – Jul. 2025

Platform Leviathan: NVDIA A100s/H100s Infrastructure for Amazon NOVA

- **GPU Lifecycle & Blacklisting System (*Cost Savings*):** Architected a scalable tracking system handling spikes of **3,000+ scaling requests**, preventing circular terminations of NVIDIA H100s/A100s on EC2 and projecting to save **$1.5 million annually**.
- **Automated Reliability Orchestrator (*Efficiency*):** Designed a "Bad GPU" identification workflow for a fleet of **7,000+ GPUs**, reducing node troubleshooting time by **90%** (from 10 hours to 1 hour) and saving **~100 engineer-hours per month** via automated isolating strategies.
- **API Optimization & Centralized Monitoring (*Performance*):** Re-engineered DynamoDB access patterns to improve API response times by **70%** and deployed a centralized monitoring framework that achieved a **100% system failure detection rate** for business-critical incidents

SKILLS

- ***Programming Languages****: (Golang), Python, C/C++, Java, TypeScript, SQL.*
- ***AI Infrastructure & GPU:*** *NVIDIA A100/H100 Optimization, Ray (Cluster/Serve), DCGM, AWS EFA, PyTorch Distributed.*
- ***Cloud Native & Kubernetes:*** *Kubernetes Internals (Operators, CRDs, Virtual Kubelet), AWS EKS, Helm, Docker, Containerd.*
- ***Infrastructure as Code (IaC):*** *Terraform, AWS CDK, AWS CloudFormation, Ansible.*
- ***Observability & Networking:*** *Prometheus, Grafana, eBPF, gRPC, TCP/IP, VPC Networking.*

PUBLICATIONS

- Hsu, Yan-Cheng; Li,Y.H.; Chang,C.C.; Harfiya, Latifa N. 2020. "*Generalized Deep Neural Network Model for Cuffless Blood Pressure Estimation with Photoplethysmogram Signal Only.*" *Sensors* 20, no. 19: 5668.
- Latifa N.; Hsu, Yan-Cheng; Li, Y.H.; Wang, J.C. 2023. *"On the Optimal Self-Supervised Multi-Fault Detector for Temperature Sensor Data" APSIPA ASC 2023* (Oral Presentation)
- Amazon AGI et al. 2024. "The Amazon Nova family of models: Technical report and model card"