

Craig Cooper

Senior Agentic AI Lead | Houston, TX | 713-705-1556 | craigcooper87911@gmail.com | [linkedin.com/in/craig-c-51b405146](https://www.linkedin.com/in/craig-c-51b405146)

Summary

Senior AI engineering leader with 9+ years architecting and scaling production agentic systems and cloud-native infrastructure for IBM TechZone. Led two consecutive platform GA launches in H1 2026: AskTZ V2, a multi-agent assistant, and ContextForge, an enterprise MCP gateway, spanning agent architecture, MLOps, AI safety, and GPU-accelerated inference on Kubernetes/OpenShift. Combines hands-on LangGraph/MCP engineering with team leadership (6-8 engineers), security remediation, and VP-level platform strategy.

Skills

AI / ML	LangGraph, MCP Architecture, watsonx.ai, Granite Models, LoRA Fine-Tuning, RAG, Granite Guardian (AI Safety / Guardrails)
Languages	Python, Go, TypeScript / JavaScript, NestJS, Bash, YAML
Infrastructure	Red Hat OpenShift, Kubernetes, Tekton, ArgoCD, GitOps, NVIDIA MIG / H100 GPU
Cloud & Data	IBM Cloud, Kafka (Event Streams), Milvus, Cloud Object Storage, PostgreSQL, HashiCorp Vault
Observability	OpenTelemetry, Grafana Tempo, Loki, Prometheus, IBM Concert

Experience

Senior Agentic AI Lead | IBM April 2024 - Present | Houston, TX (Remote)

AskTZ Platform: Led a 6-8 engineer team building and operating AskTZ, IBM TechZone's production multi-agent AI assistant across Reservation Creation, Extensions, and Documentation workflows; drove Beta to GA (completed March 2026) across two major releases, totaling 103 resolved engineering tickets, now processing ~800 requests and serving ~200 unique users weekly.

MCP Gateway: Designed and shipped ContextForge, an enterprise-grade MCP control plane integrating 15+ tools across provisioning, billing, and support domains with full OIDC/SSO auth middleware, tool scoping, and Agent-to-Agent (A2A) protocol support for cross-agent orchestration; later deployed a shared instance with IBM Concert integration for cross-team observability.

RAG Pipeline: Built AskTZ's Documentation retrieval pipeline — Drupal content ingestion, semantic chunking, and Milvus vector search exposed as MCP tools through ContextForge — with a grading/reranking layer to reduce hallucination and improve answer relevance.

AI Safety: Deployed Granite Guardian 3.3-8b for input/output screening, PII detection, and RAG sanitization, live across all four production pipeline functions on OpenShift AI.

Observability: Built an OTEL pipeline (otelcol-contrib to Grafana Tempo + IBM COS) delivering full trace visibility across ContextForge and auth services; implemented agent quality and NPS tracking via MLflow.

Platform Strategy: Led a migration feasibility assessment for TechZone's Cirrus platform, presenting three critical architectural blockers to VP-level stakeholders to align leadership on platform roadmap priorities and risk mitigation.

Security: Leading an active pentest remediation cycle across TechZone, closing token-exposure findings and strengthening the platform's security posture.

Software Development Manager | IBM May 2022 - April 2024 | Houston, TX (Remote)

Platform Services: Led design and development of Golang, Python, and PHP microservices powering IBM TechZone's provisioning and deployment tooling on OpenShift.

Kubernetes Automation: Built a Kubernetes Operator (Go) automating OpenShift cluster bootstrapping for IBM Software Deployment, reducing manual provisioning errors.

CI/CD & Compliance: Designed Tekton CI/CD pipelines automating IBM Cloud Pak installations across OpenShift; remediated CVEs across six clusters, led HashiCorp Vault secret rotation, and conducted RBAC audits.

Senior Software Developer / Tech Lead | IBM April 2019 - May 2022 | Houston, TX (Remote)

Microservices & CI/CD: Delivered full-stack microservices and automation across IBM Cloud Paks deployment tooling on OpenShift, building foundational CI/CD and platform engineering capabilities.

GitOps & Platform Reliability: Built an ArgoCD + Tekton GitOps pipeline across multiple OCP clusters, established a CP4I Event Streams / Keycloak integration, and resolved a critical platform availability outage, hardening the environment against recurrence.

Education

B.S. Mathematics | University of Texas at San Antonio | 2018