Nikhil Kaul

Seattle, WA · (919) 884-1542 · nkaul2345@gmail.com · linkedin.com/in/nikhil-kaul5 · nikhilkaul.com

SUMMARY

Technical product manager building AI infrastructure and agentic platforms for Kubernetes. Led the public launch of the seven-skill Azure/AKS-Skills open-source suite across GitHub Copilot, Claude Code, Azure SRE Agent, and compatible hosts. Shipped Azure Chaos Studio's ground-up platform rewrite, cutting serving COGS 90%, expanding capacity 100×, and saving $3M+ annually; now leads capacity-efficiency strategy across Microsoft's 200M+ core fleet. Hands-on builder known for cross-org influence across engineering, capacity, support, and finance.

EXPERIENCE

Microsoft — Azure Cloud Native (AKS/VMSS) · Product Manager *Redmond, WA · Jul 2025 – Present*

- Led the public launch of Azure/AKS-Skills, shipping seven open-source, read-only-first capabilities for AKS troubleshooting, cost optimization, GPU/inference operations, network capture, readiness, and cluster design across GitHub Copilot, Claude Code, Azure SRE Agent, and compatible hosts.
- Lead capacity-efficiency strategy across Microsoft's 200M+ core fleet; built the outcome-based measurement model and drove cross-org AKS/VMSS product direction for workload right-sizing, mixed-SKU placement, and Karpenter-based provisioning.
- Shipped the Azure Chaos Studio ground-up platform rewrite to public launch in 9 months: cut serving COGS 90% ($300K → $30K/month), saved $3M+ annually, and expanded capacity 100× (1K → 100K customers) at the same fully-loaded cost; own the roadmap for 30+ engineers across the US and India.
- Defined product and safety requirements for enterprise agentic support: resource-scoped evidence access, customer-output controls, auditability, and human approval.

Microsoft — Azure Observability · Product Manager, Azure Chaos Studio *Redmond, WA · Apr 2023 – Jul 2025*

- Took over Azure Chaos Studio and led the product turnaround: grew to 200K+ fault-injection experiments per year with the world's largest financial institutions as production users, and drove the ground-up redesign that shipped as v2.
- Championed agentic development, inner/outer-loop tooling, and hermetic run-to-green testing across the team — feature throughput rose ~30× YoY while headcount fell 70%, and new fault development compressed from 3 months to 3 days.
- Shipped Chaos Agent (health signals, fault debugging, auto-upgrade), private networking, CMEK, Customer Lockbox, and Terraform modules; 50+ Fortune 500 resilience engagements; keynote demos at Microsoft Ignite 2023 and Build 2024.

Microsoft — Windows + Devices · Product Manager *Redmond, WA · Aug 2021 – Apr 2023*

- Joined Windows Fundamentals; promoted within one year. Led arm64/x64 manufacturing parity, enabling top OEMs (HP, ASUS, Acer, Lenovo) to build Windows-on-Arm devices on their existing x64 imaging pipelines; owned Windows Optional Features (20M+ MAU).
- Pioneered firmware-based OS recovery over Wi-Fi: drove Wi-Fi driver adoption into UEFI across four silicon vendors (Qualcomm, MediaTek, Realtek, Intel), compressed under 2MB to fit 16/32MB firmware — now shipping across major OEM devices; authored the technical whitepaper (modeled $250M support-cost reduction). (PM intern, 2020: browser-based Windows installation, presented at CVP level.)

PROJECTS & OPEN SOURCE

Agency Co-Work (2025) — Built a desktop workspace for agentic AI that ships with its source and a built-in terminal, enabling users to fork and extend personal copies.

Microsoft Hackathon, 1st Place (2025) — MCP-based resilience scanner for AI-generated apps; won the global executive challenge under Azure CEO Scott Guthrie.

Pedi QuikCalc — Top-20 iOS health app: 10,000+ paid downloads ($10); used by Doctors Without Borders teams in Africa and US pediatric clinics for child dosing.

EDUCATION

Duke University — B.S. Computer Science & Philosophy (double major) *Durham, NC · 2017 – 2021*

3.78 GPA · 2330 SAT · Dean's List · Duke Data+ research: ML plagiarism detection across literature and film.

SKILLS

Infrastructure: Kubernetes (AKS, Karpenter/node auto-provisioning, GPU/inference operations, scheduling and bin packing), Azure, VMSS, Terraform, Docker; capacity management, utilization and cost-to-serve analytics, demand forecasting, quota systems, SLOs, incident response.

Engineering & AI: Python, TypeScript/JavaScript, Swift, React, Node.js, Electron; GitHub Copilot, Claude Code, Azure SRE Agent, MCP; hands-on open-source agent capability development.

Product & leadership: 0→1 platform products, roadmap ownership, A/B testing and funnel analysis, executive communication, cross-org programs and influence without authority.