Shwetal Mehta

Ashburn, Virginia (open to relocation) · mehta.shwetal@gmail.com
linkedin.com/in/shwetal-m · github.com/shwetalmehta72 · shwetalmehta.com (case studies and live demos)

Hands-On AI Engineering Leader | Self-Hosted Open-Weight LLMs, Agentic Systems, ML in Production | Built and Ran a $200M+ Technology-Services P&L

I set the strategy, stand up the infrastructure, build the first systems myself, then hire and lead the team that scales them. At NovaSystems I design, build and operate production AI: a self-hosted open-weight LLM behind a cost-capped routing layer, agentic decision systems with fail-closed controls, a billion-row real-time data platform and ML models trained on cloud GPUs. At CenturyLink and Lumen I built a $200M+ ARR cloud, hosting, application and security managed-services business with full P&L, owned the capacity economics of dedicated SAP HANA infrastructure, and prototyped generative AI that went into production. AI and ML in production since 2017.

CORE COMPETENCIES

AI: LLMs (OpenAI, Anthropic, Google, open-weight) · self-hosted models (llama.cpp) · agentic AI, custom Claude Skills and MCP servers · RAG · prompt engineering (one-/few-shot, hallucination guards) · machine learning (PyTorch, gradient boosting) · AI governance and operating controls

Technical: IT service management and AIOps (ServiceNow, IPsoft, Ansible) · hybrid infrastructure and cloud (AWS, Azure, Google Cloud) · SAP HANA platforms · enterprise applications and end-user services · multi-vendor SD-WAN and network-as-a-service · managed security services · real-time data platforms

Business: P&L and business-unit leadership · productization and service catalog design · pricing and deal structuring · enterprise pursuits ($50M+ TCV) · ITIL service management · global onshore/offshore delivery · capacity planning and vendor management · executive client relationships

EXPERIENCE

NovaSystems — Founder & Chief Architect · Ashburn, VA — Aug 2025 – present

- My self-funded systematic trading and AI systems business, trading my own capital (no outside customers or investors). I design, build and operate its systems hands-on with AI coding tools (Claude Code, Cursor, Windsurf) against written specifications.
- Self-hosted open-weight LLM (gpt-oss-120b on llama.cpp) for routine, high-volume inference, behind a routing layer I designed: vendor and frontier models (OpenAI, Anthropic, Google, Kimi) handle work that needs stronger judgment, and a daily spend cap falls back to the local model for routine tasks.
- Agentic decision systems in daily production with fail-closed controls: human-in-the-loop and emergency-stop modes, approval gates before consequential actions, a risk governor. Custom Claude Skills and MCP servers; prompt engineering with one-shot and few-shot patterns and hallucination guards (code computes every number; the model ranks and explains).
- Real-time data platform: live market-data streaming over WebSockets (675 symbols), 30+ scheduled feeds with freshness checks, 1B+ rows in TimescaleDB and ClickHouse, FastAPI services.
- Machine learning in production: PyTorch transformer and XGBoost/LightGBM/CatBoost models trained on cloud GPUs (Modal, earlier Google Vertex AI), Optuna tuning, evaluation on held-out data before promotion.

CenturyLink → Lumen Technologies · Washington DC area — Oct 2012 – Jul 2025

Senior Director, Product Management — Managed Services (Lumen, Oct 2020 – Jul 2025) · Director (May 2017 – Oct 2020) · Principal, Remote Infrastructure & Application Managed Services (May 2014 – May 2017) · Client Engagement Manager (Oct 2012 – May 2014)

- Brought in by the SVP of Professional Services to build a managed-services business from under $1M of staff augmentation and one-off projects. Created the service catalog, pricing models, ITIL-aligned delivery, SLAs, SLOs and KPIs, and standard contracts; grew it to $200M+ ARR by 2025 and $1B+ in lifetime contract value with full P&L. CenturyLink CEO's Award for Innovation (2015).
- SD-WAN “Rosetta Stone” (2024): my idea and prototype. A cross-vendor command and configuration library for Versa, Viptela and Meraki covers ~80% of repeatable work; OpenAI GPT generates vendor-specific commands for engineers to review and apply. My operations team built the production version; tier 1–2 resolution times fell 50%.
- AI pricing and SOW engine (2024–25): architected and built a working prototype on Lumen Private Cloud. A deterministic catalog and pricing model compute every number, retrieval supplies service definitions and pricing data, and Anthropic and OpenAI models only suggest the next question, never seeing customer names. Evaluated by ~20 solutions engineers on 200 live opportunities and back-tested on 200 historical deals: SOW preparation fell from 40+ hours to 2–3, with no legal rework.
- Owned the P&L and capacity economics of managed SAP HANA on dedicated large-memory infrastructure in our data centers: demand forecast against purchase, utilization, and redeployment as customers scaled.

- Rebuilt delivery twice: 100% onshore (2014) to 80% offshore (2018) to automation. Productized automated triage and resolution on IPsoft, ServiceNow and Ansible (2018; sold as a paid add-on from 2020): by 2019, across the accounts measured, ~87% of ~8,000 monthly tickets were touched by automation and 1,800+ hours a month of manual effort removed. Defined machine-learning ticket-classification use cases with the data-science team (2017).
- Jointly led the business with the operations leader over a 500+ person delivery organization (U.S., India, Singapore, London, Poland): I owned the offer, pricing and P&L. Designed and owned the 50-person operations command center (KPIs, runbooks, budget). Direct team of 5 product managers and 4 solution specialists; set direction for 15 sales engineers.
- Led every pursuit above $50M in total contract value, including three multi-service-line contracts of $100M+ in manufacturing and consumer goods; the pursuits I led produced ~$100M of the $200M run-rate. 200+ customers across manufacturing, consumer goods, retail and healthcare.

Infosys — Practice Engagement Manager, Infrastructure & Cloud Services Dec 2011 – Oct 2012

- Account manager for cloud and infrastructure solutions in banking; accounts included five of the ten largest U.S. banks at the time.

HCL Technologies — Client Partner, Mid-Atlantic · Washington DC area Jul 2008 – Nov 2011

- About six accounts, mostly in consumer goods and manufacturing, with an annual book of business of ~$30M.

Cyberwerx — Director of Operations (also CFO) · Cary, NC and Bangalore Nov 2001 – 2007

- Built the company's offshore development center in Bangalore from scratch (relocating from the U.S. to do it) to 100 engineers serving six clients, with full P&L, hiring, legal and compliance responsibility, and took it to CMMI Level 5. Clients included a global networking-equipment maker (technical-assistance center and software lifecycle) and a telecom-equipment maker.
- Led the sale of the India center end to end: buyer meetings, business case, sales package, contracting and post-sale transition.

Pliant Systems (formerly Broadband Technologies) — TAC Manager 1998 – 2001

- Led an 8-engineer technical-assistance center, presales and the demo center for one of the first fiber triple-play companies; field-deployed with four of the largest U.S. carriers.

Earlier: Consultant, Unified Network Systems (network management), 1997–98 · Principal Consultant, Imonics: on-site implementation of document capture and workflow for a European mobile carrier, a medical-billing operation and a hospital system, 1995–96.

EDUCATION · CERTIFICATIONS · AWARDS

BSE, Electrical Engineering and Biomedical Engineering (double major), Duke University, 1994; coursework in AI (expert systems, pattern recognition) · MBA, UNC Kenan-Flagler Business School, 2001

Certifications: Machine Learning Foundations (University of Washington, 2016) · Data Science (Johns Hopkins, 2016) · AI Strategy and Governance (University of Pennsylvania, 2025) · Agentic AI and AI Agents for Leaders (Vanderbilt, 2025)

Awards: CenturyLink CEO's Award for Innovation (2015) · President's Club · multiple Sales Achievement Awards (CenturyLink / Lumen, 2015–2022)

Languages: English · Hindi · Spanish (limited working)