



Prem Moola

Enterprise AI platform builder | Governed AI & agentic systems | Data platform modernization | Engineering transformation | High-trust technical leadership

prem@premmoola.com | linkedin.com/in/premmoola | www.premmoola.com

EXECUTIVE VALUE PROPOSITION

Enterprise AI and platform executive helping regulated organizations operationalize governed AI, modernize data platforms, and scale engineering execution.

Technology executive with 20+ years of experience across financial services, SaaS, and media, combining executive ownership with deep technical credibility in enterprise AI platforms, data modernization, cloud architecture, and global engineering leadership. Strong fit for senior AI platform, data, engineering, and enterprise transformation roles that require turning platform strategy into business outcomes, including 80% infrastructure cost reduction, 75% operational cost reduction, 40% reporting-cycle improvement, 70% faster time-to-insight, 5x deployment-velocity improvement, and 99.99% uptime.

EXECUTIVE HIGHLIGHTS

20+ years leading enterprise platform and engineering transformation

Built governed AI platforms for regulated and security-sensitive environments

Reduced infrastructure cost by 80% and operational cost by 75%

Improved reporting cycle time by 40% and time-to-insight by 70%

Scaled global engineering teams across enterprise and founder-led environments

Former Goldman Sachs VP and BNY Mellon Senior Director; builder of Srasta.ai and Diggt Product Studio

CORE COMPETENCIES

Enterprise AI platform strategy; governed AI infrastructure; LLM infrastructure and RAG; agentic AI systems; model routing and policy-aware tool execution; data engineering platforms; cloud-native architecture; platform modernization; distributed systems; regulatory and trading data platforms; auditability, lineage, and recovery controls; engineering leadership; vendor strategy; operating-model transformation.

PROFESSIONAL EXPERIENCE

Srasta.ai

Builder and Platform Lead

December 2025 - Present

- Built Srasta.ai, a customer-controlled enterprise AI platform for regulated environments, unifying governed inference, retrieval, identity-aware access, auditability, and operator-managed deployment in one system.

- Designed the platform's trust boundary so customer documents, prompts, responses, identity data, audit records, backups, and runtime state stay inside the customer's own perimeter unless an operator explicitly opts in to sharing them.
- Architected the AI execution layer: OpenAI-compatible inference and RAG APIs, LiteLLM-based routing, vLLM- and MLX-backed local inference, Milvus retrieval, MinIO-backed storage, and governed tool execution.
- Built a gateway-centered control plane handling authentication, model-level RBAC, tenant and role enforcement, rate limiting, signed principal propagation, service proxying, audit logging, and fail-closed access by default.
- Implemented governed retrieval and ingestion: tenant isolation, ACL-aware document access, provenance tracking, grounding prompts, audit streaming, and permission-aware retrieval.
- Led Srasta Membrane, the governed memory and state boundary for AI execution, covering approvals, lineage, drift evaluation, memory lifecycle, rollback, runtime proof, and cell-pinning evidence.
- Defined install, onboarding, and day-2 operating patterns across Docker/Kubernetes-style deployments: identity bootstrap, admin workflows, Cloudflare Tunnel exposure, runtime health checks, recovery controls, and operator trust requirements.

Diggt

Builder and Product Studio Lead

August 2025 - Present

- Built Diggt as a mobile-first product studio and shared platform layer for SaaS products across property operations, workforce management, personal productivity, subscriptions, analytics, CRM, and payments.
- Established Studio Platform as the reusable foundation across live products, standardizing authentication, notifications, analytics, CRM, deployments, release operations, billing, and subscription capabilities to reduce incremental build effort.
- Embedded small language model (SLM) patterns directly into the iOS and Android UI layer, moving task assistance, summarization, recommendations, and context-aware interactions closer to the mobile experience instead of treating AI as a remote-only backend feature.
- Designed on-device and edge-assisted AI interaction patterns for lower-latency mobile workflows, privacy-sensitive local context, resilient UX under variable network conditions, and AI-native product differentiation across the portfolio.
- Launched SureLease as a cross-market rental property management SaaS for US and India workflows, including owner and tenant operations, onboarding, lease/rent workflows, subscriptions, notifications, and mobile/web delivery. Website: surelease.diggt.cloud.
- Launched Shyftly as a workforce-management SaaS platform for small-business owners, managers, and employees, then prepared beta onboarding with a 100-location QSR owner using the shared Diggt operating model. Website: shyftly.diggt.cloud.
- Launched Habit as a live mobile-first habit and goals product using a local-first interaction model with secure backup, reminders, analytics, subscriptions, and AI-assisted user experience patterns. Website: habit.diggt.cloud.

- Operationalized a low-cost shared cloud platform for product delivery, messaging, subscriptions, analytics, and release operations while keeping product experimentation fast and infrastructure spend disciplined.

BNY Mellon

Senior Director, Alternatives Data

September 2024 - July 2025

- Directed enterprise alternatives data strategy and modernization of AI-native data platforms across a \$500B AUM asset-services data estate.
- Led the Alternatives Data team across three time zones, owning architecture and strategy for approximately 1,500 funds, six accounting systems, and approximately 2,000 product/client-facing operational users.
- Defined enterprise Alternatives Data Strategy aligned to compliance, platform modernization, reporting quality, and business priorities.
- Unified ingestion from six accounting systems into a governed schema and operating model, reducing custom-reporting cycle time by 40% from requirements gathering through final client submission.
- Operationalized governed AI query and report-building workflows, reducing business-user time-to-insight by 70%.
- Institutionalized audit-controlled reporting pipelines where every report action and change was captured for traceability.
- Won the BNY AI Hackathon with an anomaly-detection solution that was adopted into production monitoring.

Third Summit / Altheon.io

Vice President of Engineering

March 2021 - January 2024

- Led engineering for an enterprise media SaaS platform serving professional creative workflows and major media organizations including CNN and Disney.
- Scaled the globally distributed engineering organization from 2 to 45 while supporting 2 enterprise customers and approximately 3,000 platform users across US and UK workflows.
- Re-architected the initial MVP into a microservices-based enterprise media platform with approximately 15 services managing approximately 800TB of data and multi-region transcoding clusters that auto-tuned capacity based on request volume.
- Reduced infrastructure costs by 80% by replacing an external cloud transcoding dependency with an in-house cloud-based transcoding service, plus architecture optimization and vendor renegotiation.
- Maintained 99.99% uptime for more than a year across pipelines and infrastructure supporting global customers and demanding production workloads.
- Integrated platform workflows with Adobe, Final Cut Pro, and DaVinci Resolve, reducing editorial turnaround time by 50%.
- Moved release operations from ad hoc delivery to structured, certified, well-tested biweekly releases.
- Helped drive \$8M ARR through camera-to-cloud partnerships, workflow integrations, and strategic ecosystem delivery.

- Supported NFT creation and submission to Ethereum using smart contracts as part of platform expansion.

CAT Technology

Vice President of Engineering

November 2017 - March 2020

- Led enterprise engineering modernization across private cloud, analytics, business-process digitization, and DevOps automation.
- Modernized infrastructure strategy by moving dev/QA workloads to private cloud, reducing monthly hosting spend from approximately \$30K to approximately \$8K while keeping production cloud-hosted.
- Established distributed analytics and data-processing platforms for enterprise reporting and operational workloads.
- Increased deployment velocity 5x, enabling weekly releases across multiple client platforms through SDLC standardization, agile operating practices, CI/CD automation, and disciplined QA/recovery workflows.
- Led a 40-member engineering organization across 20% onsite and 80% offshore delivery, supporting approximately 80 clients.
- Digitized business processes across HR, sales, engineering, marketing, timesheets, CRM, and service delivery operations.
- Owned OpEx/CapEx planning, third-party vendor relationships, negotiations, and business/technology alignment across modernization programs.

Goldman Sachs

Vice President, Engineering

September 2004 - May 2017

- Led engineering strategy and platform development for global trading analytics, regulatory reporting, data platforms, entitlement systems, and high-volume financial workflows.
- Directed global trading analytics, regulatory reporting, entitlement, and enterprise data-platform engineering across high-volume financial workflows.
- Managed a 46-node core analytics cluster with live-live regional isolation supporting fixed income and global business workflows across New York and London.
- Supported high-volume financial data platforms processing roughly 2TB of new data daily across trading analytics, regulatory reporting, warehouse, and entitlement workflows.
- Reduced platform operational overhead by 75% over six months through workflow redesign, automation, issue-pattern analysis, and self-service adoption across new teams.
- Created a scalable entitlement framework applying logical access rules across enterprise data platforms and governed access workflows.
- Established platform analytics and automation capabilities that improved performance diagnosis, cluster operations, deployment speed, and upgrade reliability across 16 production clusters.
- Modernized strategic data-ingestion and warehouse-readiness frameworks to support curated enterprise consumption and regulatory reporting workloads.
- Transformed OTC derivative regulatory-reporting architecture across Credit, Rates, FX, Equity, and Commodities, supporting daily and near-real-time reporting for global regulators.

- Improved regulatory-platform resilience and maintainability by redesigning partitioning, workflow isolation, connectivity, and reconciliation patterns across core reporting systems.
- Reduced Tier A support overhead through federated development and validation models for critical gateway workflows.
- Delivered front-office trading and sales technology including trade-entry tools, institutional sales merchandising, and electronic trading platform enablement.
- Served as Co-COO for Technology Asian Professionals Network, partnering with HCM and technology/business leadership on mentoring, retention, career development, budgeting, and divisional programming.

FX Solutions, Citigroup, and Cantor Fitzgerald

Software Engineer / Financial Technology Roles

2001 - 2004

- Built financial technology systems before joining Goldman Sachs.

EDUCATION

Fairleigh Dickinson University

MS, Computer Science

University of Madras

BS, Computer Science Engineering

TECHNICAL SKILLS

Enterprise AI and Data Platforms: enterprise AI platforms, governed AI infrastructure, LLM infrastructure, RAG, model routing, agentic workflows, regulated data platforms, distributed analytics, auditability, lineage, retrieval systems, and production evaluation loops.

Cloud and Platform Engineering: cloud-native architecture, platform automation, Kubernetes and container operations, serverless delivery, developer platforms, observability, release engineering, cost optimization, resilience, and recovery controls.

Software and Security: API design, authentication and authorization, RBAC, OIDC/JWT, policy-aware workflows, microservices, service proxying, mobile/web delivery, and security hardening for regulated environments.

Leadership and Transformation: global engineering organizations, executive communication, operating-model design, vendor strategy, OpEx/CapEx planning, product strategy, and business-aligned platform modernization.