

## THE UNIFIED DIGITAL FRAMEWORK

The modern enterprise faces a critical paradox: while individual technologies evolve at breakneck speed, integrating these tools often lags, creating fragmented systems. Our approach dissolves these silos. By integrating Data Science, Full Stack Development, DevOps, and QA into a singular "Unified Digital Framework," we transform experimental technology into production-grade assets that drive measurable ROI for your organization. In the current digital economy, competitive advantage is no longer determined by access to technology alone, but by an organization's ability to scale its implementation. Businesses have been investing excessively in AI, cloud platforms, modern applications, and automation, yet many struggle to convert these investments into tangible business outcomes.

As technological ecosystems expand in digital areas, organizations face growing complexities. The fragmented data sources, inconsistent data quality, and time-consuming manual analysis made it difficult for Data science teams to handle. Fullstack development teams struggled with applications tied to tightly coupled systems and complex integrations. The manual infrastructure management, environmental inconsistencies, and error-prone deployment process put pressure on the DevOps teams. Meanwhile, QA teams grappled with difficulty in validating expanding applications across various platforms and devices, often dependent on manual testing, leading to slower launches. Together these challenges create severe bottlenecks demanding solutions that not only solve individual technical problems but also connect workflows across disciplines to create a more unified, scalable, and resilient digital ecosystem.

Endure works towards building cutting-edge systems that are involved in providing end-to-end solutions to all the possible bottlenecks by helping to build and scale digital products across Data Science, Full-Stack Engineering, DevOps, and Quality Assurance. The Data Science service focuses on building intelligent AI systems, retrieval-augmented generation (RAG) workflows, evaluation frameworks, and autonomous monitoring solutions that improve the accuracy, scalability, and reliability of AI-driven experiences. Modern full-stack development involves the company building high-performance web applications, enterprise platforms, dashboards, and multi-brand ecosystems using modern technologies and scalable software architectures on both sides- frontend and backend. In addition, Endure's DevOps services aim to establish reliable, resilient, and cloud-native infrastructure through automation, Kubernetes orchestration, observability, and disaster recovery strategies. Quality Assurance ensures product reliability through reliable automated testing that validates both expected user journeys and edge cases, ensuring faster releases, reduced risk, and consistent high-quality user experiences. By bridging the gap between innovation and execution, Endure aims to enable organizations to transform technology investments into expandable and measurable business outcomes.

## **Retrieval-Augmented Generation (RAG) & Knowledge Graph Systems**

Designing and developing Retrieval-Augmented Generation (RAG) pipelines to enable context-rich AI interactions by integrating semantic retrieval systems, vector embeddings, APIs, MCP integrations, and multiple structured and unstructured data sources. The work involves building scalable retrieval workflows that improve the quality and relevance of AI-generated responses across enterprise environments. Focus areas include semantic search, embedding generation, chunking strategies, prompt engineering, retrieval optimization, and improving context handling while reducing hallucinations in production AI systems.

Additional responsibilities include integrating agent-based workflows and retrieval pipelines to support dynamic and context-aware AI applications capable of handling enterprise-level use cases and data interactions.

Working extensively with Knowledge Graphs (KGs) and graph-based retrieval systems using technologies such as Neo4j, Dgraph, Kuzu, and Cognee to improve structured knowledge retrieval and relationship-aware querying. Responsibilities include designing graph schemas and relationships, enabling dynamic and multi-hop querying, implementing relationship traversal logic, and improving graph-based semantic retrieval for enhanced contextual understanding. The work also includes handling graph consistency challenges, optimizing query workflows, and improving the way structured relationships are interpreted within AI systems.

Additionally, involved in debugging and stabilizing AI systems in production environments, including resolving workflow failures, retrieval inconsistencies, API issues, worker startup problems, and graph database limitations. Experience also includes deployment support, end-to-end testing, validation of AI workflows, and collaborating with cross-functional teams to improve reliability, scalability, and overall performance of enterprise AI applications. The overall focus is on building scalable, reliable, and context-aware AI systems capable of supporting real-world business and knowledge-driven use cases.

## **Evaluation Maturity: Building Reliable LLM-as-a-Judge Frameworks with DeepEval and MLflow**

Designing and developing scalable evaluation frameworks for GenAI applications using LLM-as-a-Judge methodologies, DeepEval-style evaluation metrics, and MLflow-based experiment tracking systems. The work involves building reliable evaluation pipelines for both RAG and non-RAG workflows by integrating custom metrics, automated response validation, experiment comparison systems, and evaluation microservices within enterprise AI platforms. Focus areas include semantic similarity evaluation, hallucination detection, completeness scoring, response precision, token overlap metrics, evaluation schema design, and improving the consistency and reliability of AI-generated outputs across enterprise use cases. Additional responsibilities include integrating evaluation workflows with AgentFlow pipelines, AMP microservices, LiteLLM routing, and scalable backend architectures to support production-ready GenAI evaluation systems.

Working extensively on experiment tracking, evaluation persistence, and model comparison workflows using MLflow and AMP-integrated storage systems to improve evaluation maturity and observability. Responsibilities include implementing asynchronous MLflow logging services, experiment/run tracking, metadata persistence, S3 mirroring, evaluation artifact management, and backend integrations for scalable evaluation infrastructure. The work also includes designing run comparison frameworks with metric-wise analysis, baseline vs. comparison evaluation, LLM filtering, and elapsed-time tracking to help users identify the best-performing GenAI configurations more effectively.

## **Agentic Workflows & MCPs**

Developing autonomous monitoring agents using orchestration platforms like Flowise, with a focus on integrating multiple observability tools through Model Context Protocols (MCP). The work involves connecting Kubernetes, Grafana, and ClickHouse MCPs with the agent to give it complete visibility across development and production environments. When something breaks in dev or prod, most developers do not know where to start and end up checking random logs, switching between tools, and spending a lot of time just trying to make sense of raw data before even concluding. The agent addresses this directly: instead of manually sifting through logs and metrics across disconnected systems, users can simply ask in natural language about a failing resource or run a routine health check. The agent pulls pod statuses and cluster events from Kubernetes, fetches real-time Prometheus metrics through Grafana, and runs batch log analysis on ClickHouse to identify patterns and anomalies, then surfaces a clear, prioritized summary along with commands to fix the issue. This cuts through the noise significantly, getting the user to an answer in seconds so they can start fixing immediately.

Worked with agent development frameworks like AWS Strands to build and configure agents equipped with the right set of tools for handling multi-step tasks. The work includes integrating vector databases into the agent pipeline, with attention to selecting appropriate chunking strategies that balance retrieval accuracy and semantic relevance. Additional focus areas include building evaluation datasets to systematically assess agent response quality, covering both accuracy and reasoning across different types of queries. The overall effort spans tool integration, retrieval pipeline setup, and response evaluation aimed at making agents reliable enough for consistent, real-world use.

## **Agentic AI Systems, MCP Integrations & Evaluation Frameworks**

Designing and developing agentic AI systems by building scalable workflows that leverage MCP integrations, multi-agent orchestration, and tool-driven architectures. The work focuses on creating intelligent agents capable of dynamic decision-making, interacting with external systems, and solving complex enterprise use cases. This includes developing agents using frameworks such as LangChain, CrewAI, Strands, and Flowise, along with implementing interactive user experiences through A2UI for dynamic input handling. Emphasis is placed on robust agent design, prompt engineering, and improving workflow reliability and adaptability in production environments.

Working extensively on building a comprehensive evaluation system for agentic AI workflows, including MCP-based agent evaluations and evaluation microservices. This involves implementing LLM-as-a-judge methodologies combined with custom code-based metrics to assess performance, accuracy, and consistency of agent outputs. Leveraging frameworks such as DeepEval, the work includes designing evaluation pipelines, creating high-quality datasets for benchmarking, and enabling systematic validation of agent workflows. Additionally, worked on the LLM benchmarking system to identify optimal models based on system requirements such as latency, cost, and response quality.

Additional work includes retrieval-augmented and knowledge-driven agent systems leveraging vector and graph databases to enhance contextual understanding and response accuracy. Contributions include chunking strategies, optimizing retrieval workflows, and improving context handling within agent pipelines. The work also involves MCP package creation and integration through internal artifact repositories, enabling reusable and standardized components across AI systems. Overall, the focus is on building reliable, scalable, and well-evaluated agentic AI systems tailored for enterprise applications.