

LLMOps for Scalable and Reliable AI

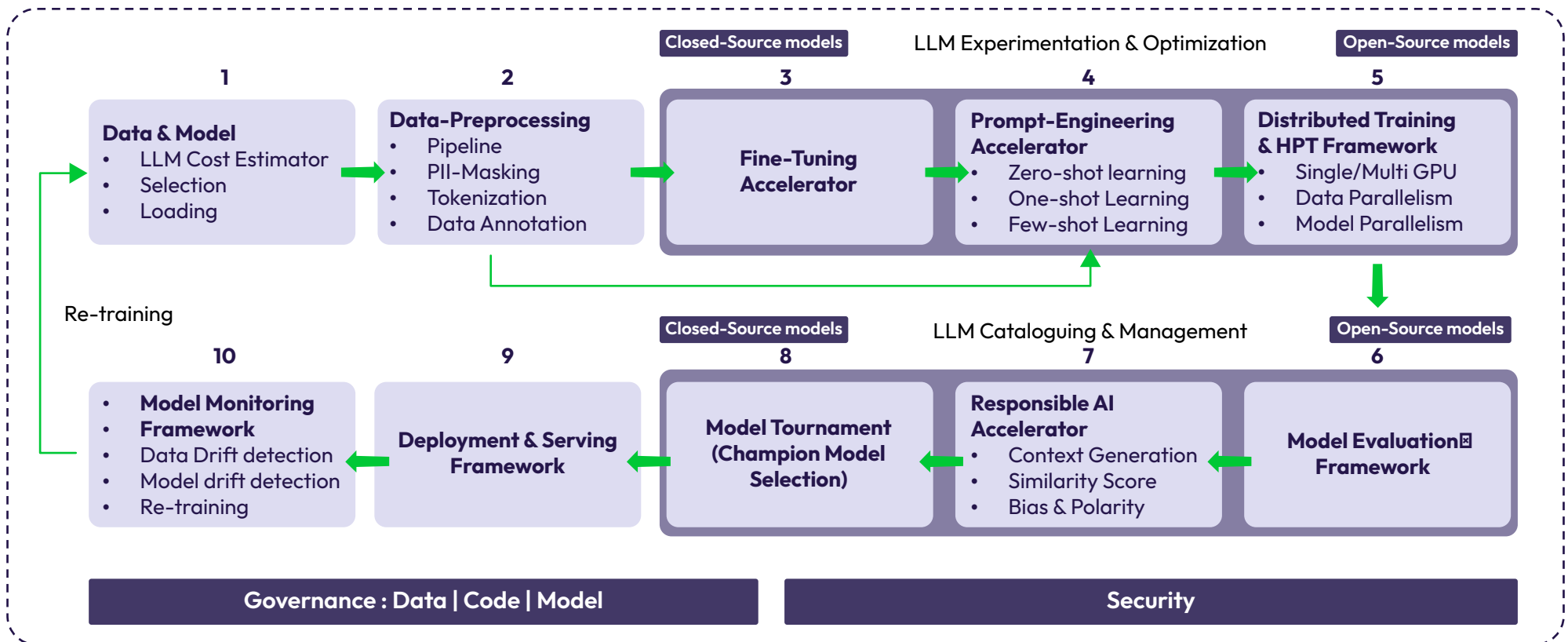
Brillio's LLMOps suite accelerates your LLM journey across model selection, cost estimation, data preprocessing, fine-tuning, and deployment.



Enterprises face significant challenges in deploying LLMOps, starting with the steep technical learning curve required to translate business requirements into functional LLM-based solutions. Building both the code and infrastructure is complex, often resulting in fragmented efforts across teams. This disjointed approach leads to inefficiencies, as multiple teams “reinvent the wheel” rather than leveraging a unified, centralized strategy. Additionally, the absence of a common data source for raw data, cleansed data, or fine-tuned models creates bottlenecks. Despite these hurdles, LLMOps is essential for enterprises seeking to deploy AI models that are scalable, reliable, and consistently aligned with business goals.

Brillio's LLM Ops Suite to Accelerate Your LLM Journey

Brillio's LLM suite enables enterprises to accelerate their LLM journey by streamlining every step, from model selection and cost estimation to data preprocessing, fine-tuning, and deployment. With tools for PII-masking, tokenization, and data annotation, the suite ensures data is ready for LLM applications. Fine-tuning and prompt-engineering accelerators enhance learning capabilities, while distributed training optimizes model performance across GPUs. Real-time model monitoring, champion model selection, and responsible AI tools improve accuracy and reliability, supporting telecom providers in deploying scalable, compliant, and responsive LLM-based solutions effectively.



Why Brillio: Our Suite of Accelerators

Fine-Tuning Accelerator

This accelerator refines and customizes foundational models to task-specific datasets, using parameter-efficient techniques, low-code templates, and hyperparameter tuning, reducing training time and data requirements for cost-effective fine-tuning.

Prompt Engineering Accelerator

This tool optimizes language models for task-specific needs through effective prompting, using low-code templates that minimize training iterations and data requirements, offering cost-efficient model customization.

Inferencing Accelerator

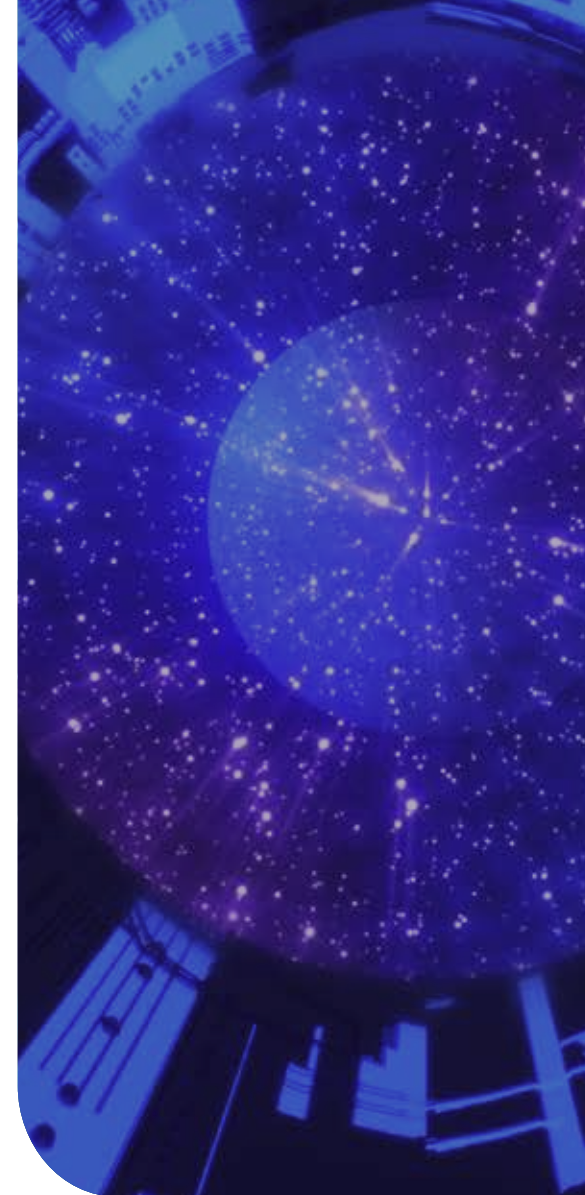
Enables fast, cost-effective predictions under high traffic using quantization, mixed-precision, and multi-GPU inferencing, reducing latency and optimizing performance.

Intelligent Model Health Monitoring Accelerator: Automatically detects data or model drift, initiates retraining when needed, and incorporates feedback, providing productivity gains through automated retraining schedules.

Responsible AI Accelerator

Enhances model clarity and explainability by tracking prediction lineage, similarity scores, and bias/polarity detection, reducing hallucinations in predictions.

LLM Cost Estimator: Facilitates planning and budgeting by estimating LLM costs, analyzing scalability across cloud platforms, and offering customizable cost projections.



Responsible AI: Brillio's Governance Structure

Brillio's Responsible AI governance framework is built on a foundation of ethical, transparent, and compliant AI practices, structured around four core pillars:

1. Data-centric

Emphasizes data ethics, regulatory compliance, and lineage reliability, with a focus on justness, privacy, explainable AI (XAI), and ensuring data quality through rigorous checks.

2. Code-centric

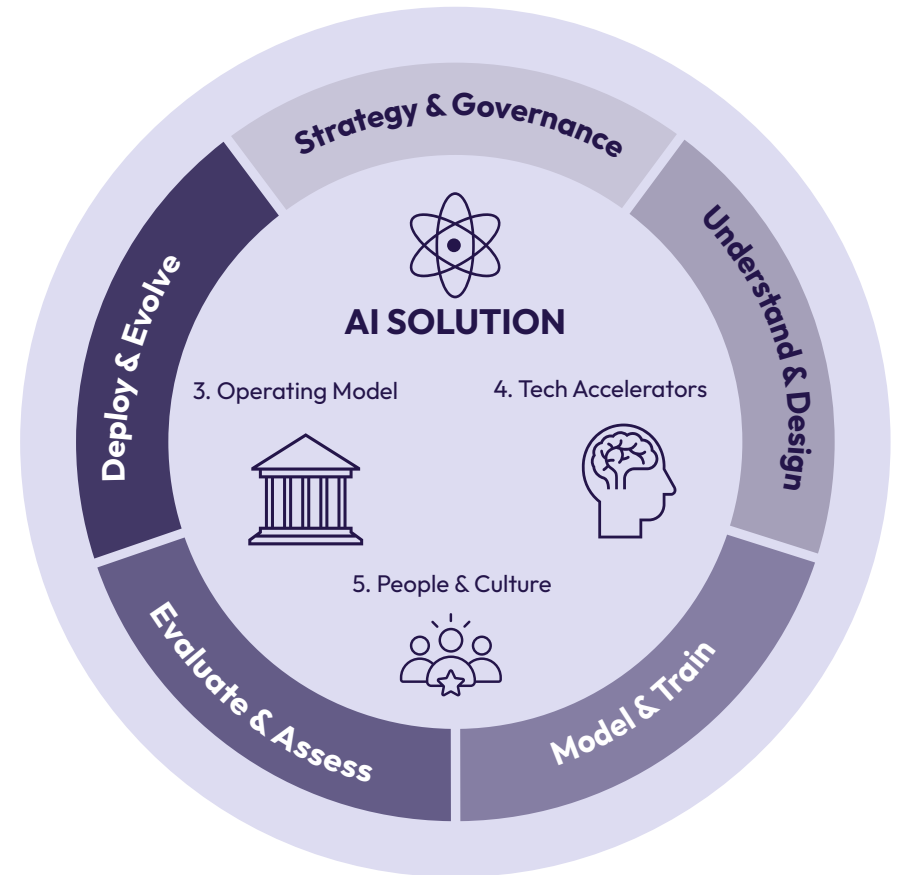
Ensures high standards in code quality, accountability, and reusability, reinforcing clarity and reliability across the AI development lifecycle.

3. Model-centric

Focuses on bias detection, edge-case testing, model ownership, comprehensive audits, and ongoing evaluation to mitigate risks and ensure output quality.

4. Domain-centric

Incorporates domain-specific expertise to ensure compliance with legal, ethical, and regulatory policies relevant to each industry.



LLM Gateway: A Centralized Application for a Telecom Giant

A Finnish multinational telecommunications and electronics company sought a centralized application for seamless LLM development and deployment, catering to diverse user personas like citizen data scientists, data scientists, and business stakeholders. The client faced challenges with non-standardized LLM processes, persona-specific issues in application management, and release management difficulties for LLM models.

Brillio developed an LLM gateway application, creating a unified solution to help different personas (citizen data scientists, data scientists, and business stakeholders) with building, deploying, and maintaining. Brillio enhanced the customer's platform to have a one-stop-solution for all activities for both open and closed source.

Key features

- Low-code/No-code (LC/NC) solutions with templated code
- Standardized pipeline and code for prompt engineering and fine-tuning
- Integration of LLM model deployment in LLM Gateway using Azure OpenAI
- Use of accelerators to template pipelines and code for efficient prompt engineering and fine-tuning

Client Impact

- Faster onboarding for model experimentation
- Increased team enablement through intuitive UI
- Enhanced access management and scalability

ABOUT BRILLIO

Brillio is one of the fastest growing digital technology service providers and the partner of choice for many Fortune 1000 companies seeking to turn disruptions into competitive advantages through innovative digital adoption. We help clients harness the transformative potential of the four superpowers of technology: cloud computing, Internet of Things (IoT), artificial intelligence (AI) and mobility. Born digital in 2014, we apply our expertise in customer experience solutions, data analytics and AI, digital infrastructure and security, and platform and product engineering to help clients quickly innovate for growth, create digital products, build service platforms, and drive smarter, data-driven performance. With 17 locations across the U.S., the UK, Romania, Canada, Mexico and India, our growing global workforce of nearly 6,000 Brillians blends the latest technology and design thinking with digital fluency to solve complex business problems and drive competitive differentiation for our clients. Brillio was certified by Great Place to Work in 2021, 2022 and 2023.



<https://www.brillio.com/>

Contact Us: info@brillio.com

