OPEA Overview¶

OPEA (Open Platform for Enterprise AI) is a framework that enables the creation and evaluation of open, multi-provider, robust, and composable generative AI (GenAI) solutions. It harnesses the best innovations across the ecosystem while keeping enterprise-level needs front and center.

OPEA simplifies the implementation of enterprise-grade composite GenAI solutions, starting with a focus on Retrieval Augmented Generative AI (RAG). The platform is designed to facilitate efficient integration of secure, performant, and cost-effective GenAI workflows into business systems and manage its deployments, leading to quicker GenAI adoption and business value.

The OPEA platform includes:

- Detailed framework of composable microservices building blocks for state-of-the-art GenAI systems including LLMs, data stores, and prompt engines

- Architectural blueprints of retrieval-augmented GenAI component stack structure and end-to-end workflows

- Multiple micro- and megaservices to get your GenAI into production and deployed

- A four-step assessment for grading GenAI systems around performance, features, trustworthiness and enterprise-grade readiness

OPEA Project Architecture¶

OPEA uses microservices to create high-quality GenAI applications for enterprises, simplifying the scaling and deployment process for production. These microservices leverage a service composer that assembles them into a megaservice thereby creating real-world Enterprise AI applications.

Microservices: Flexible and Scalable Architecture¶

The GenAI Microservices documentation describes a suite of microservices. Each microservice is designed to perform a specific function or task within the application architecture. By breaking down the system into these smaller, self-contained services, microservices promote modularity, flexibility, and scalability. This modular approach allows developers to independently develop, deploy, and scale individual components of the application, making it easier to maintain and evolve over time. All of the microservices are containerized, allowing cloud native deployment.

Megaservices: A Comprehensive Solution¶

Megaservices are higher-level architectural constructs composed of one or more microservices. Unlike individual microservices, which focus on specific tasks or functions, a megaservice orchestrates multiple microservices to deliver a comprehensive solution. Megaservices encapsulate complex business logic and workflow orchestration, coordinating the interactions between various microservices to fulfill specific application requirements. This approach enables the creation of modular yet integrated applications. You can find a collection of use case-based applications in the GenAI Examples documentation

Gateways: Customized Access to Mega- and Microservices¶

The Gateway serves as the interface for users to access a megaservice, providing customized access based on user requirements. It acts as the entry point for incoming requests, routing them to the appropriate microservices within the megaservice architecture.

Gateways support API definition, API versioning, rate limiting, and request transformation, allowing for fine-grained control over how users interact with the underlying Microservices. By abstracting the complexity of the underlying infrastructure, Gateways provide a seamless and user-friendly experience for interacting with the Megaservice.

Next Step¶

Links to:

- Getting Started Guide

- Get Involved with the OPEA Open Source Community

- Browse the OPEA wiki, mailing lists, and working groups: https://wiki.lfaidata.foundation/display/DL/OPEA+Home

- **Open Platform for Enterprise AI (OPEA) Framework Draft Proposal**