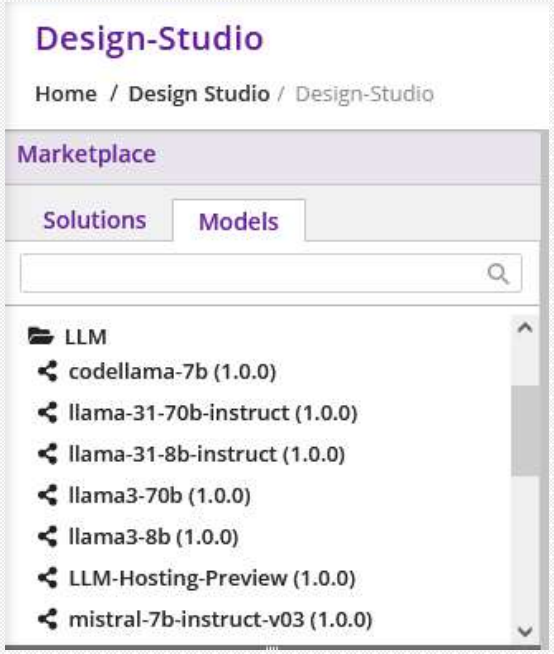
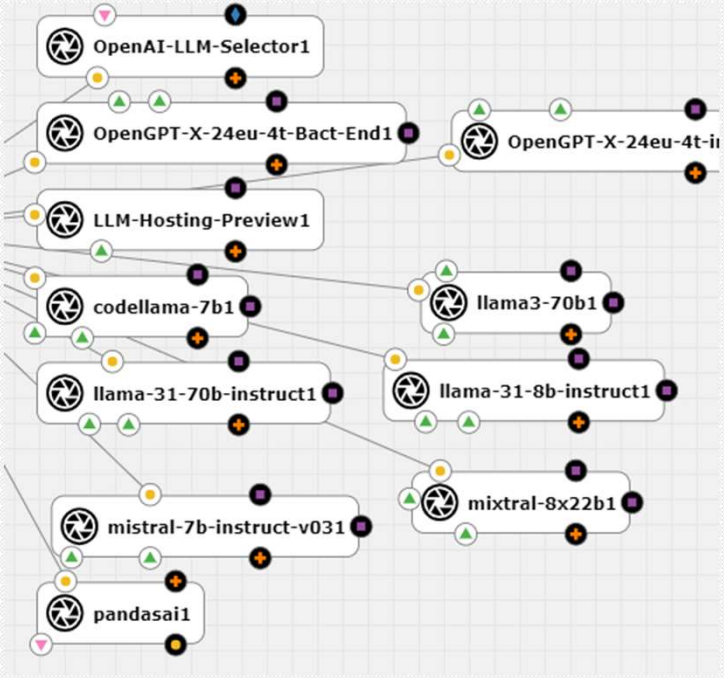
AI4Europe T4.4 Meeting - Updates

Agenda

- Unified protobuf
- Federation API
- Playground-App
- Open discussion

Overview: LLM Containers in AI-Builder

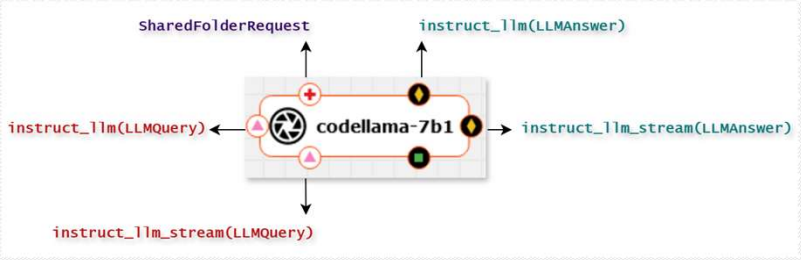




Link to Graphene Tutorial; *LLMS in AI-Builder*

The Role of the Unified Protobuf Interface

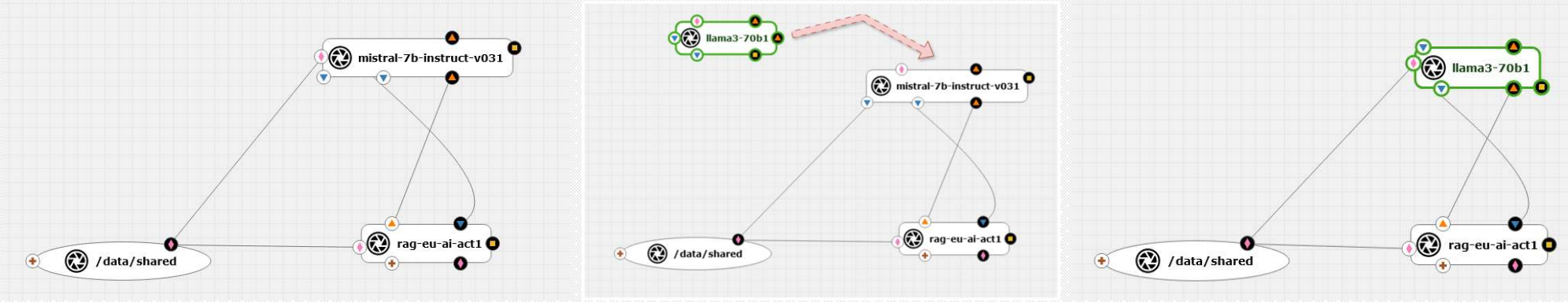
- Flexible and standardized way of defining and representing connectors for different language models in the Eclipse Graphene platform
- Offers model switching ability
- Develop language, model agnostic framework




```
syntax = "proto3";

message Empty {

}

message LLMConfig {
    double temp = 1; // Can be expanded - To be discussed.
}

message PromptInput {
    string system = 1;
    string user = 2;
    string context = 3;
    string prompt = 4; // The application-specific prompt to the LLM for processing.
}

// Encapsulates a single question string, representing a user query.
message UserQuestion {
    string question = 1;
}

// Contains a q_id field, which uniquely identifies a conversation or query. This ensures continuity and state tracking across interactions
message ConvoID {
    string q_id = 1;
}

// Bundles together the LLM configs, prompt input, user question, and conversation ID into a single message.
message LLMQuery {
    LLMConfig config = 1;
    PromptInput input = 2;
    UserQuestion qa = 3;
    ConvoID id = 4;
}

// Response from the LLM
message LLMAnswer {
    string text = 1;
    ConvoID id = 2;
}

message Status {
    string message = 1;
}

service LLMService {
    rpc instruct_llm( LLMQuery) returns( LLMAnswer);
    rpc instruct_llm_stream( stream LLMQuery) returns( stream LLMAnswer);
}
```

Current Usage

- Ability to switch between multiple LLMs with minimal efforts – Unified Protobuf Interface
- Easy to test and compare different models for various tasks.
- If a model fails or underperforms, we can alter the pipeline to a different LLM.

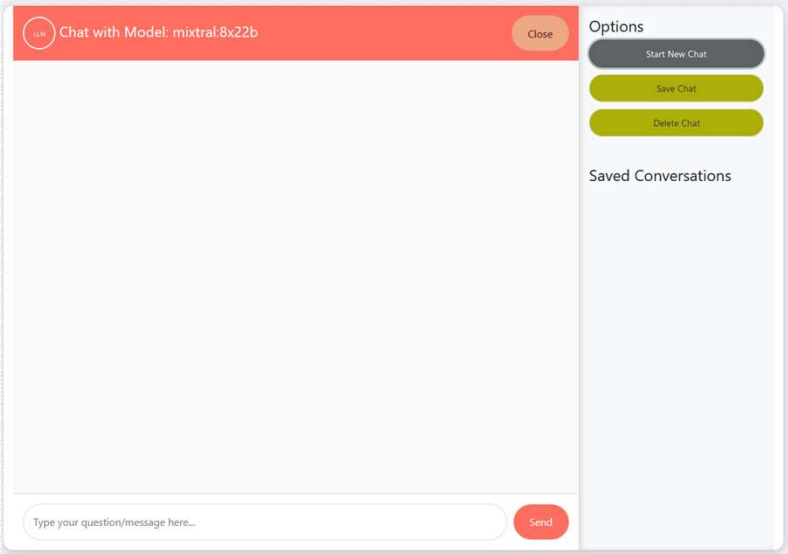


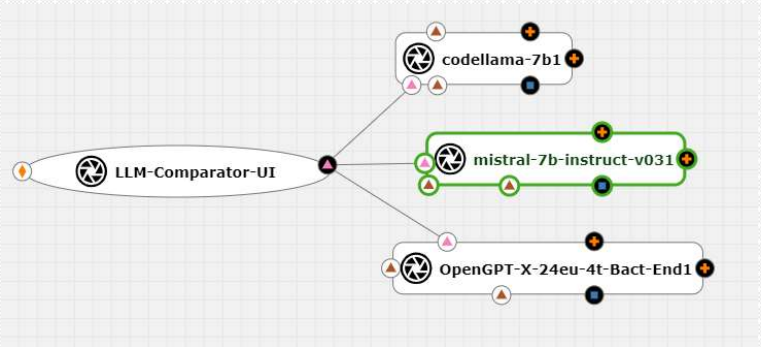
Building Blocks in a RAG EU_AI_Act Pipeline

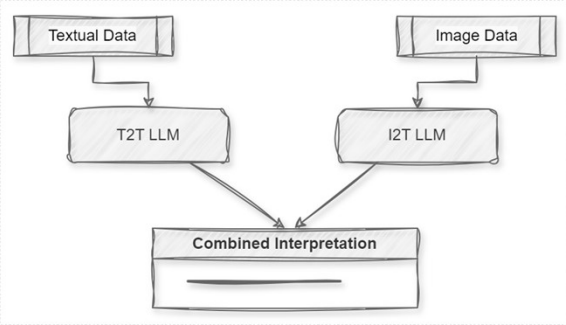
Link to Graphene Tutorial; LLMS in AI-Builder

Future Directions: Applications and Possibilities

- Serving as LLM Chatbots
- Creating custom workflows/pipelines with different modalities
- LLM-Comparator node

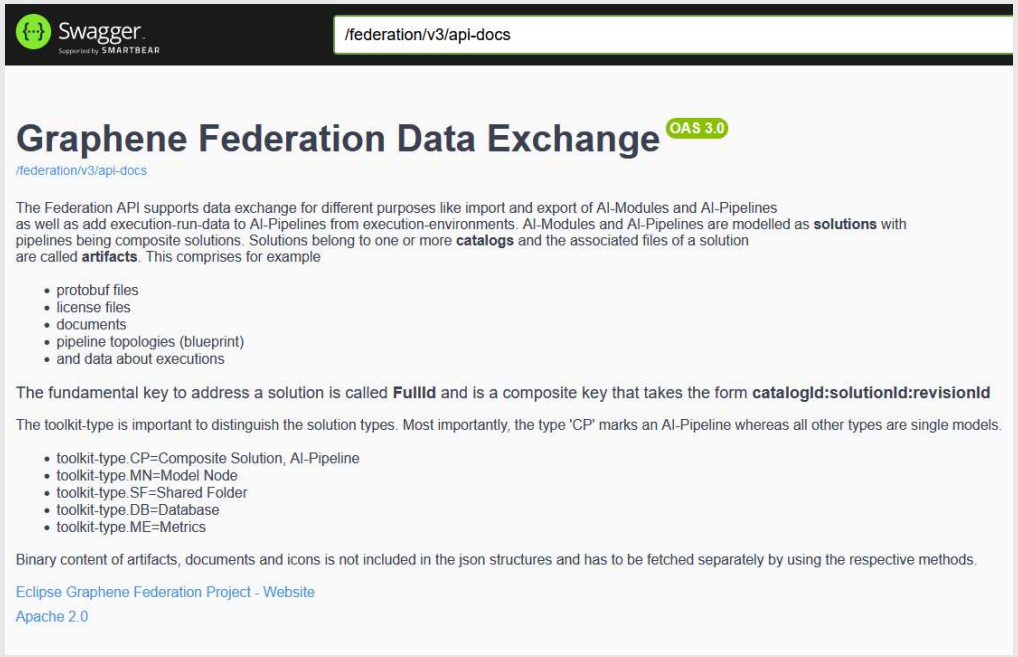


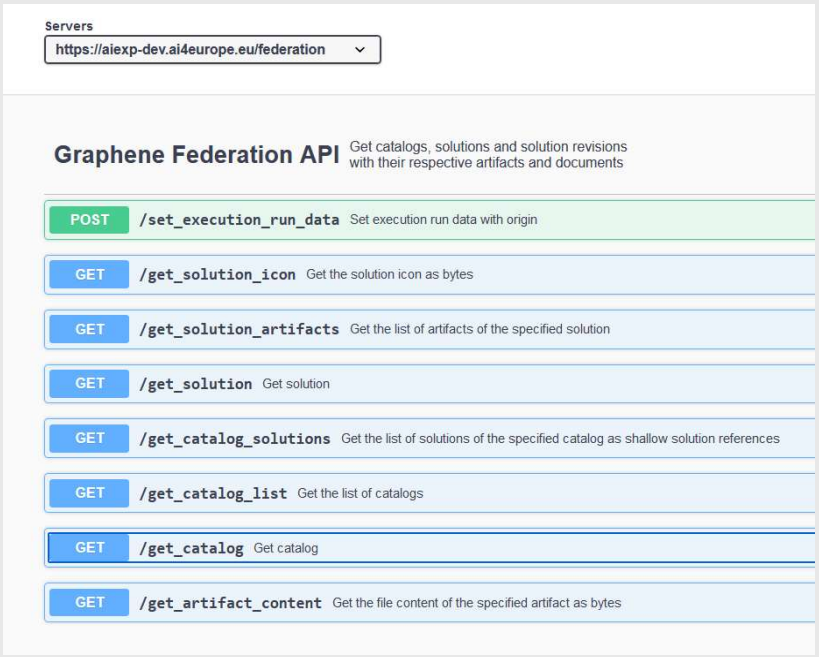




Link to Graphene Tutorial; *LLMS in AI-Builder*

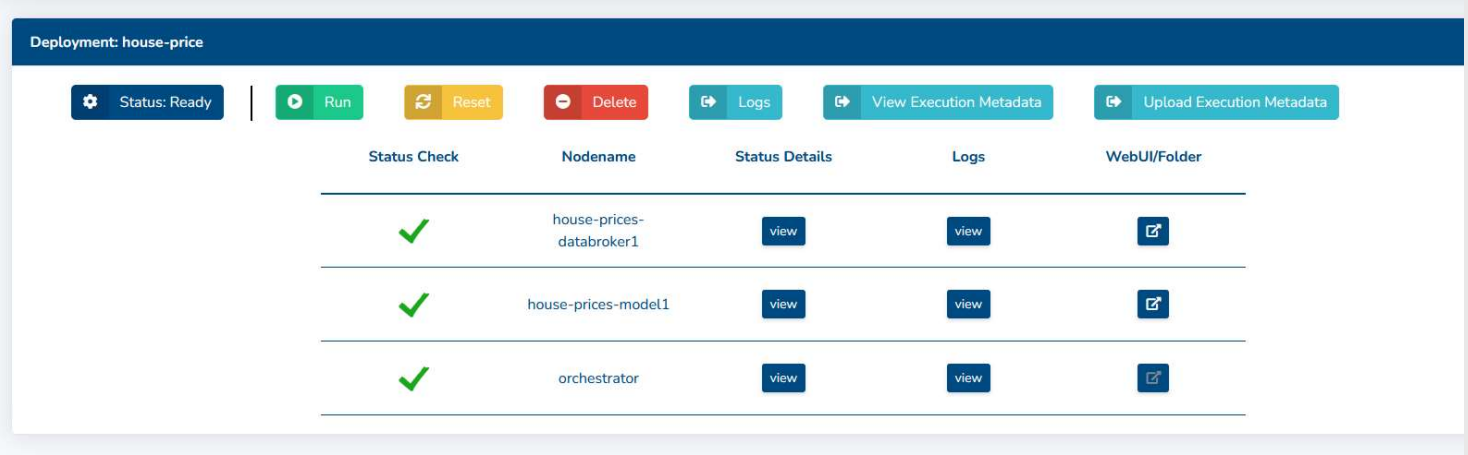
Federation API





https://aiexp-dev.ai4europe.eu/federation/swagger-ui/index.html#/

Updates to Playground-App



The playground-app has been extended to support uploading of execution-run.json to the origin graphene system of the pipeline.

Thank you