

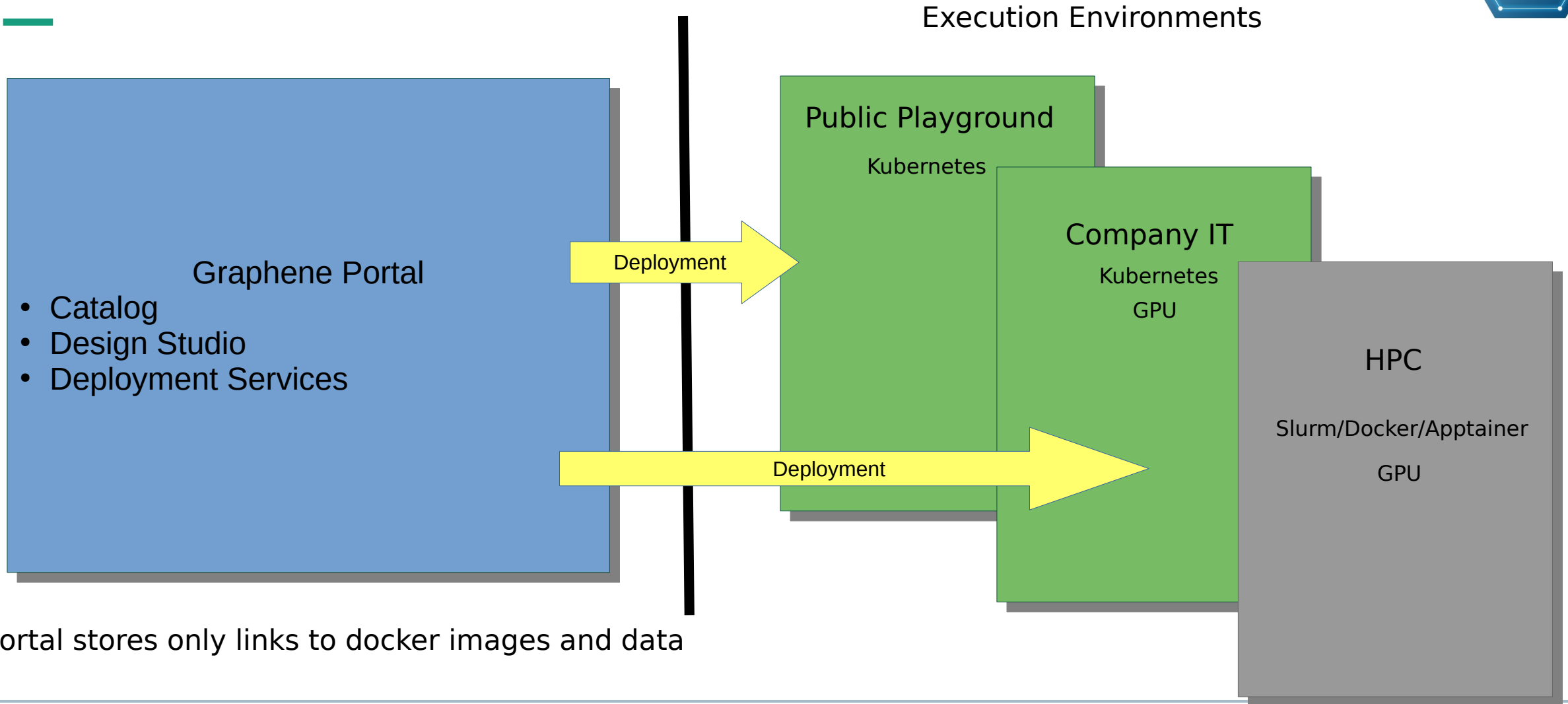


EclipseCon 2023

Applied AI with Eclipse Graphene

Martin Weiß
Project Lead Graphene

System Overview



AI Model Catalog

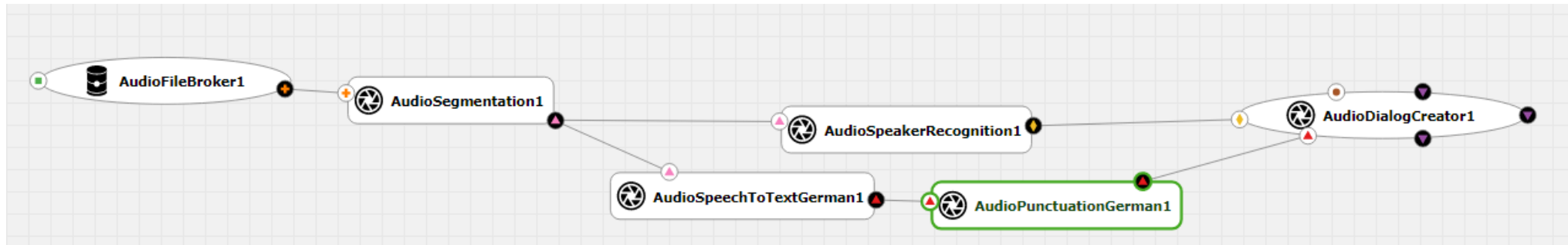
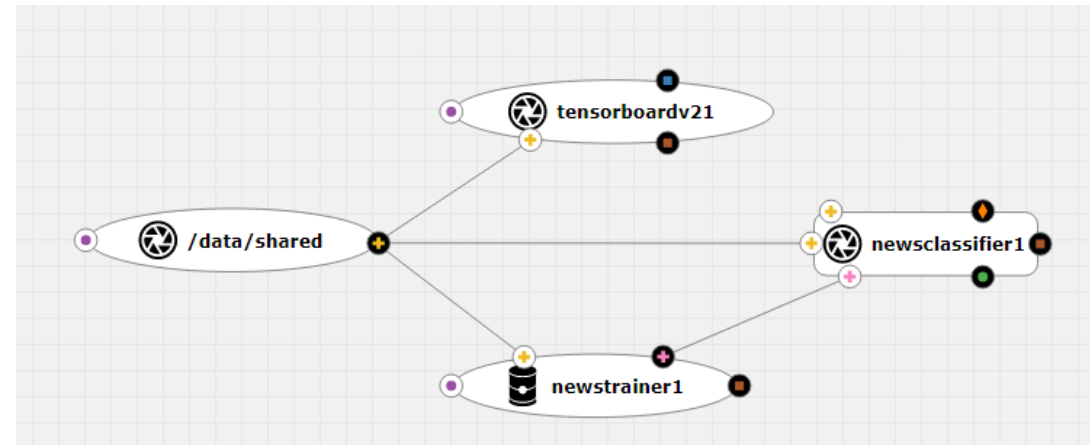
- Catalog of re-usable AI Modules
- The AI-Modules are Docker Containers
- Work in teams
- Support for commercial and open source licenses for models

The screenshot displays the 'Marketplace' interface for AI models. At the top, there are navigation options: 'All Catalogs', 'My Favorite Catalogs', and 'Select Favorite Catalogs'. Below this, the page shows 'Showing - 1 to 25 of 275 Models'. On the left side, there is a 'BROWSE BY' section with a search bar and a 'Filter By Category' section with checkboxes for Classification, Data Sources, Data Transformer, Prediction, and Regression. Below the filters, there is a 'Tags' section with various tags like 'Binary Classification', 'owl', 'Training', 'timeseries', 'electrical load forecasting', 'manufacturing', 'ai4agri', 'GUI', 'databroker', 'Tutorial', 'python', 'integrative ai', 'Air quality', 'hello world', 'Pose Estimation', 'VideoModels', 'classification', 'MultiClass Classification', 'housepriceprediction', 'connector', 'SensorThings', 'computer vision', 'Sentiment Analysis Model', 'Inspiration', 'A', 'CNN', 'covid predict', 'transformer', 'MNIST', 'Random Forests', and 'AudioMining'. The main area displays a grid of model cards. Each card includes a logo, a title, author information, a 'New' badge, a star rating, and icons for comments, views, downloads, and likes. The models shown include: 'i-nergy-load-forecasting' (Energy Load Forecasting), 'VideoObjectRecognition', 'forWoT', 'Sudoku Tutorial', 'AWDrugsModel', 'covid predict', 'AI4Agri-qualitypredictor', 'FaceAI', and 'AI4Industry Pilot Solution'. The bottom of the page features logos for 'OGC' (Making location count.), 'ITAINNOVA' (INSTITUTO TECNOLÓGICO DE ARAGÓN), and the Fraunhofer IAIS logo.

Compose AI Pipelines in the Design Studio



- Training pipeline with Tensorboard Integration
- Shared-folder concept
- Show matching models
- Application Pipeline



AI Playground



- Execution Environment
- One-Click-Deployment
- Based on Kubernetes
- Playground-App is a Graphene repo
- One-Click Update: Docker images are re-pulled on Reset

The screenshot displays the KI NRW AI Playground interface. At the top left, the KI NRW logo is visible. The main header shows the deployment name 'house-prices' and the user 'Martin'. Below the header, there are control buttons: 'Status: Ready', 'Run', 'Reset', 'Delete', and 'Logs'. A table lists the deployment components:

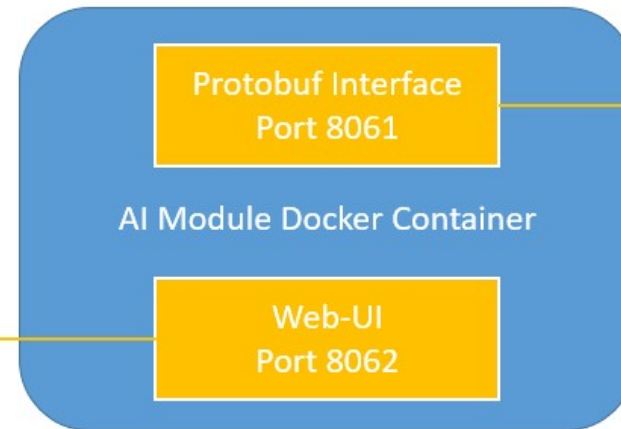
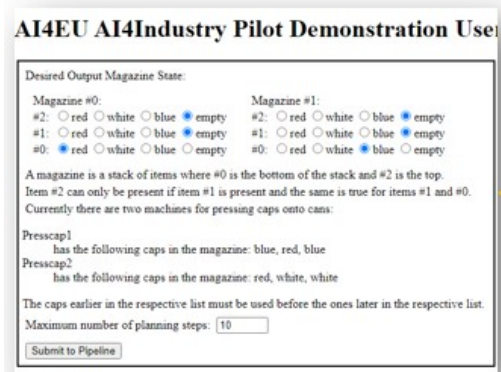
Status Check	Nodename	Status Details	Logs	WebUI/Folder
✓	hppdatabroker1	show	view	📄
✓	hppmodel1	show	view	📄
✓	orchestrator	show	view	📄

Below the table, there is a 'Solution description' section with a diagram showing two nodes: 'House-Price-Databroker1' and 'House-Prices-Prediction1'. To the right, there is an 'Overview description' section with text: 'The House Prices Pipeline is a simple example pipeline that predicts house prices. The pipeline illustrates how the price development is predicted by entering relevant parameters that provide information about the status of a property.' Below this is a 'Use case example' section: 'As an interested house owner, an estimate can be made based on the AI forecast, how much the property will increase in value or not.' At the bottom, there is a 'Usage' section.

Container Specification for composable AI-Modules



- Docker Container with additional properties
- public interface in `protobuf` / `grpc`
- optional Web-UI (highly recommended)
- No Lock-In: it continues to be a normal Docker container
- If done right, the same Docker container can be used across different execution environments (e.g. GPU or no GPU)



```
// set used version of protobuf
syntax = "proto3";

// set unique package name
package fraunhofer_demo_cpp_iris.v1;

// define input data structure
message IrisDataFrame {
  repeated double sepal_length = 1;
  repeated double sepal_width = 2;
  repeated double petal_length = 3;
  repeated double petal_width = 4;
}

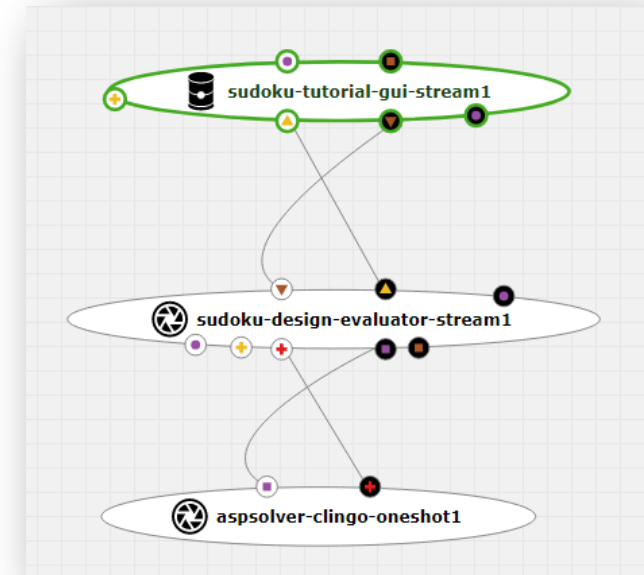
// define output data structure
message ClassifyOut {
  repeated int64 value = 1;
}

// define exposed service
service Model {
  rpc classify (IrisDataFrame) returns (ClassifyOut);
}
```

Generic Parallel Orchestrator



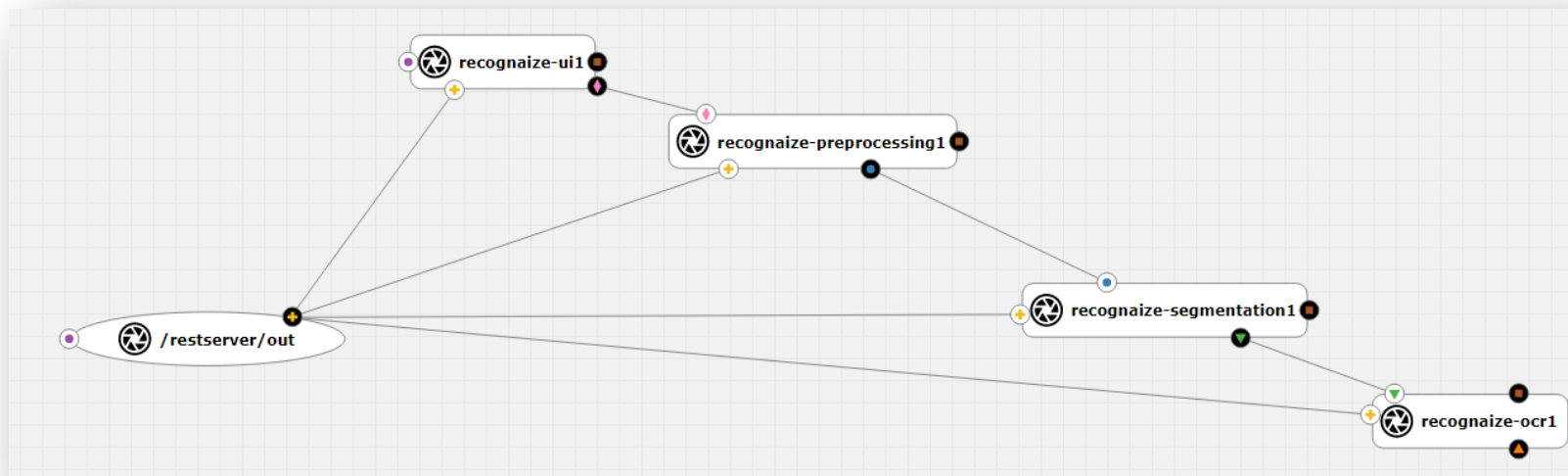
- The orchestrator executes the pipeline according to the topology file (generated by the design studio)
- The orchestrator dispatches the messages among the nodes following the edges
- Thanks to protobuf/grpc the communication stubs can be generated on the fly
- Support for grpc streaming:
 - Sensor-Input
 - UI-Input
 - Media-Streams
 - The edges can be served in parallel
- Different orchestrator implementations possible



Pipeline Definition (Topology)



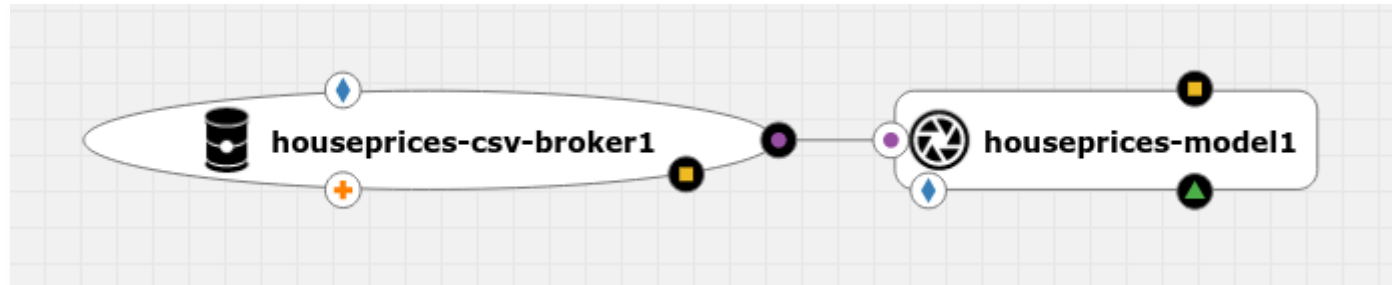
- A JSON-file which contains nodes and edges (technology neutral)
- Is generated by the Design Studio, but can be generated elsewhere (e.g. for AutoML)
- During deployment, the specific configurations for an execution environment are generated (MLOps)
- Different orchestrators are possible
- Shared Folder support



Databroker



- A Databroker is a shallow component that makes data available to the pipeline
- Usually the first node of a pipeline
- Specific for an execution environment (e.g. company IT)
- Can access the data source (credentials)
- Does not contain the dataset!
- We have already a prototype of an Eclipse Dataspace Connector



What's New in Graphene Release 1.1?



- Installation uses Kubernetes 1.26, Calico 3.26, Helm 3.12.3, Ingress-Nginx 4.7.1
- Playground-App is faster
- Support for more protobuf features:
 - enums
 - oneof
 - recursive message definitions
 - nested definitions
- Better support for directed cyclic topologies
- Jupyter-Connect
- First-Steps for Reproducibility
- And Benchmarking (Metrics)

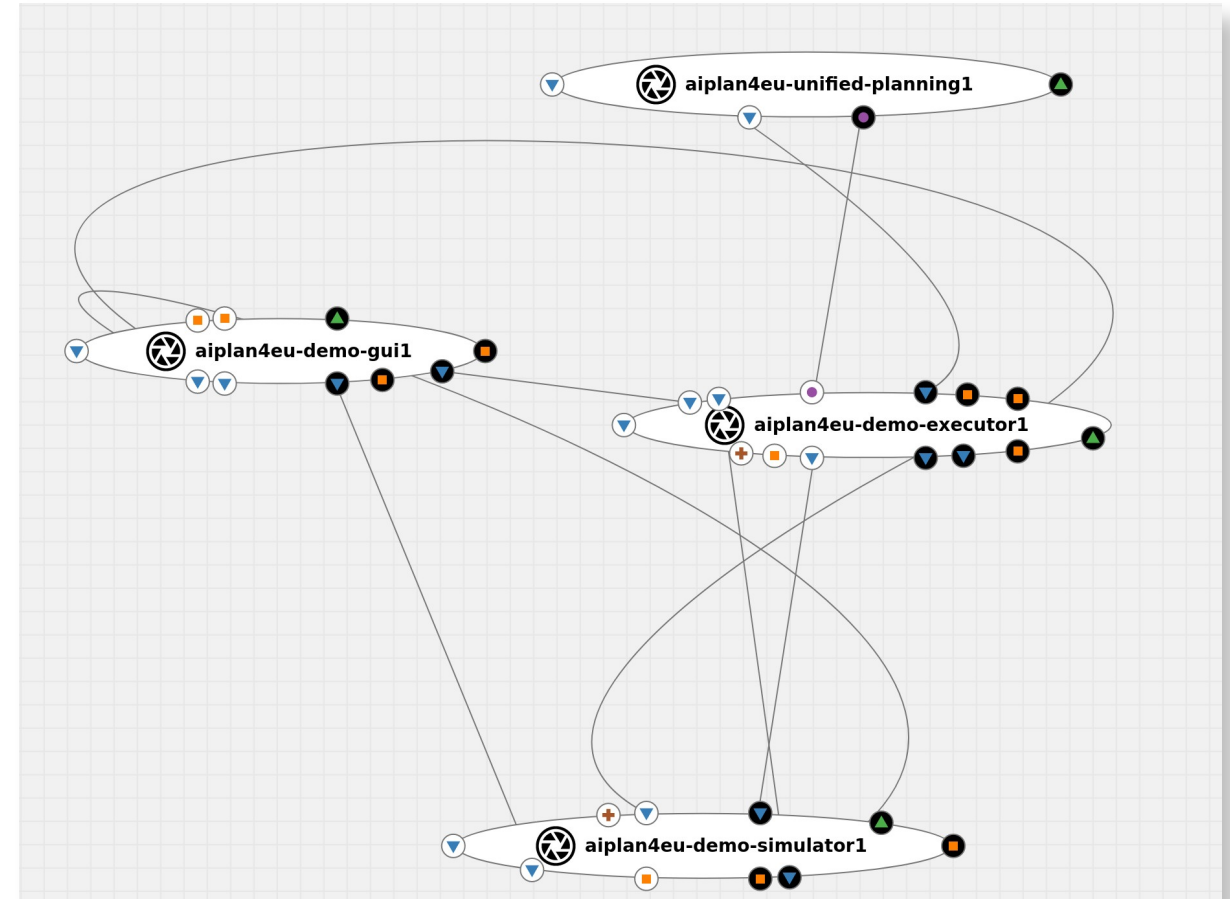
```
2 syntax = "proto3";
3
4 // As in s-expression, an Expression is either an atom or list repres
5 message Expression {
6     // If non-empty, the expression is a single atom.
7     // For instance `3`, `+`, `kitchen`, `at-robot`, ...
8     Atom atom = 1;
9     // If the `atom` field is empty, then the expression is a list of
10    // typically representing the application of some arguments to a
11    // For instance `(+ 1 3)`, `(at-robot l1)`, `(>= (battery_level) 2
12    repeated Expression list = 2;
13
14    // Type of the expression. For instance "int", "location", ...
15    string type = 3;
16    // Kind of the expression, specifying the content of the expressi
17    // This is intended to facilitate parsing of the expression.
18    ExpressionKind kind = 4;
19 }
20
```

Cognitive Architectures



- To fully leverage AI power, we need to combine LLM with other AI technologies:
 - Planning, Logic, Knowledge Graphs, Filter, Image Object Detection, OCR, Audio
 - Constant Learning, Feedback loops
 - Cognitive Architectures require directed cyclic graphs (supported by Graphene 1.1)
 - And recursive data structures (supported by Graphene 1.1)

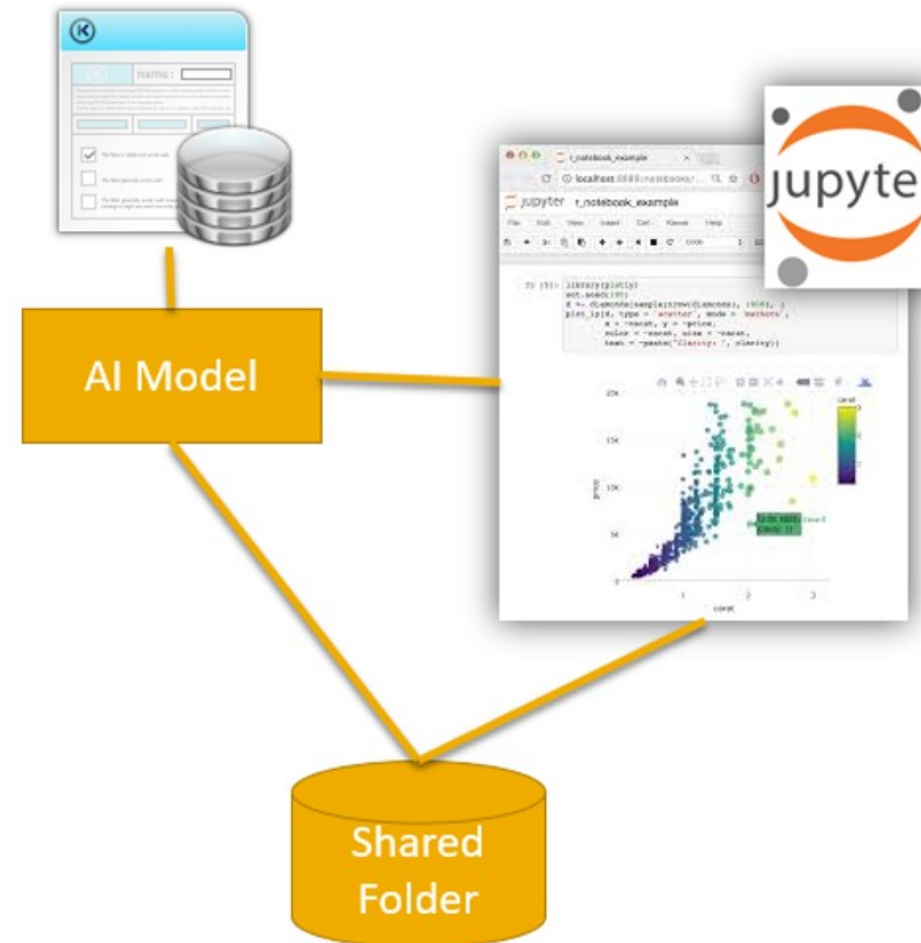
Maze-Planner



Jupyter Connect: Interactive Model Exploration



- If the user deploys a single model
- Graphene automatically generates a suitable Jupyter-Lab container
- and adds the protobuf interface definition
- and a shared folder
- for comfortable interaction with a model / AI-Module



Metadata / Benchmarking



Capture Metadata about:

- System (CPU/GPU/RAM)
- Datasets
- Docker Image Checksums
- Different Metrics
- See Container-Spec

```
"checksum": "docker-pullable://cicd.ai4eu-dev.eu:7444/tutorials/house_price_prediction/hpp_databroker@sha256:6d8c5f203a4bcc8a794b54d82a7428b36ffeb9f53b164968f576442e2bdde89f",  
"dataset_features": {  
  "type": "aiod-dataset/v1",  
  "datasetname": "The Reuters Dataset",  
  "description": "http://kdd.ics.uci.edu/databases/reuters21578/README.txt",  
  "size": "4MB",  
  "DOI_ID": "Not available"  
}
```

```
"metrics": [  
  {  
    "type": "classification-training-metrics/v1",  
    "date_time": "2023-09-07 07:28:51",  
    "accuracy": 0.9216,  
    "validation_loss": 0.9244,  
    "status_text": "success"  
  },  
  {  
    "type": "classification-testing-metrics/v1",  
    "date_time": "2023-09-07 07:28:51",  
    "F1 Score": 0.85,  
    "Specificity": 0.90,  
    "ROC-AUC": 0.92,  
    "status_text": "success"  
  }  
]
```

```
"running_time": "345s",  
"system_info": {  
  "system_name": "mwtest",  
  "fqdn": "test.playground.org",  
  "cpu": "10",  
  "gpu": "",  
  "memory": "65851412Ki"  
}
```

Challenges (Planned)



- Challenges are special types of catalogs

AI on Demand

HOME

MARKETPLACE

MY MODELS

CATALOGS

ON-BOARDING MODEL

DESIGN STUDIO

PUBLISH REQUEST

Q AND A

ML LEARNING PATH

Catalogs

Home / Catalogs

Show 10 Catalogs

CATALOG NAME	PUBLISHER NAME	SELF-PUB	ACCESS TYPE
acumos-int-fhg Internal	AI4EU Experiments	No	Restricted
AI4EU Experiments Public	acumos-int-fhg.ai4eu.eu	No	Public
Classification Challenge	DEV AI4EU Experiments	No	Public

Showing 1 to 3 of 3 Catalogs

Leaderboard (Planned)



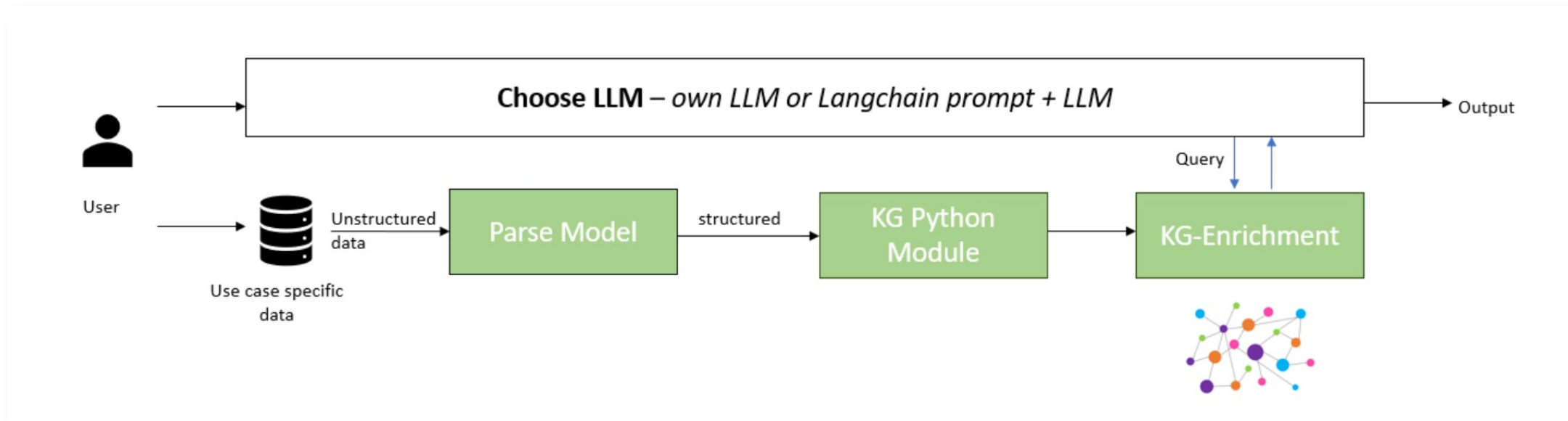
- Order based on Metrics

The screenshot displays the AI Demand Marketplace interface. The main content area shows a leaderboard for the 'Classification Challenge' catalog, listing five models ranked by metrics. The models are:

- SmartRiver** (author466, 10/13/2022): 5 stars, 0 comments, 19 views, 4 downloads.
- SWE predictor** (author1083, 10/12/2022): 5 stars, 0 comments, 18 views, 1 download.
- lexatexer-ai4hydro-proxy** (author522, 10/11/2022): 5 stars, 1 comment, 25 views, 0 downloads.
- audio-file-broker** (author207, 05/24/2022): 5 stars, 0 comments, 19 views, 2 downloads.
- AI REGIO DSS4TB** (author1198, 05/24/2022): 5 stars, 0 comments, 19 views, 2 downloads.

The interface also features a sidebar with navigation options (HOME, MARKETPLACE, MY MODELS, CATALOGS, ON-BOARDING MODEL, DESIGN STUDIO, PUBLISH REQUEST, Q AND A, ML LEARNING PATH) and a filter section for 'Filter By Category' (Classification, Data Sources, Data Transformer, Prediction, Regression) and 'Tags' (semantic web, model baseline, word2vec, artificial intelligence, anonymization, Purchasing, ai4agri, GUI, ontology, Dataset assessment, energy, ML, python, AI, gpt, language model, Pose Estimation, conditionmonitoring, classification, Forecasting, Planning under uncertainty, AI REGIO, MultiClass Classification, planning, data quality, prediction).

LLM Pipelines: Grounding LLM with Knowledge Graphs



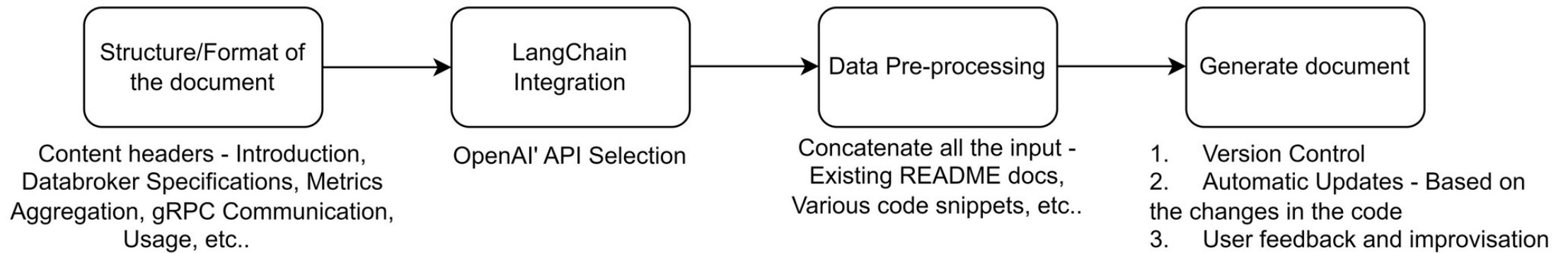
Work in progress

LLM Pipelines: Interactive Graphene Documentation



The issue aims at automatically generating documents using large language models and LangChain. By doing so, we can effectively streamline the creation of comprehensive and informative documentation.

The initial overview of the process is shown below,



Work in progress

Graphene Use Cases



- General: Building AI-pipelines from re-usable AI-modules
- AI Research: model training, benchmarking, Jupyter-Connect
- Education: practical live demonstrations and student exercises in the playground
- MLOps: simplify deployment, training pipelines, data cleaning pipelines, can support many tools
- Model showcase, catalog (internal, external)
- Application: let non ai-experts compose and deploy pipelines for their business domains
- Collaboration: work in mixed teams on PoCs
- Automated certification

Supporting Projects: Graphene Funding at least until 2028



- KI.NRW
- AI4Europe
- HumanE-AI Net
- West-AI
- Deploy-AI (starting 2024)

KI.NRW



AI4EUROPE
Supporting AIOD
2022-2025



WEST AI

HUMANE AI NET

AI on Demand
Knowledge and services
for the AI community

Useful Links



- AI4EU Experiments: <https://aiexp.ai4europe.eu/#/home>
- KI.NRW AI-Lab: <https://www.ai-lab.nrw/>
- Eclipse Graphene Project: <https://projects.eclipse.org/projects/technology.graphene>
- Eclipse Graphene Gitlab: <https://gitlab.eclipse.org/eclipse/graphene>
- Container Spec: https://gitlab.eclipse.org/eclipse/graphene/tutorials/-/tree/main/Container_Specification
- Tutorials: <https://gitlab.eclipse.org/eclipse/graphene/tutorials>

Contact

Martin Weiß
Senior AI Solution Architect
martin.welss@iais.fraunhofer.de

Fraunhofer-Institut für intelligente Analyse- und
Informationssysteme IAIS
Schloss Birlinghoven 1
53757 Sankt Augustin
Germany

www.iais.fraunhofer.de