AI4EU Experiments

Martin Welss, Fraunhofer IAIS
martin.welss@iais.fraunhofer.de

On Demand AI for everyone

AI4EU

Share, experiment
and collaborate

Sell or buy licensed
AI solutions or expertise

Try out or compose
AI solutions –
from domain user to
expert

Compatible with
established
AI toolkits

Catalog of re-usable AI Tools

Visual Pipeline Composition
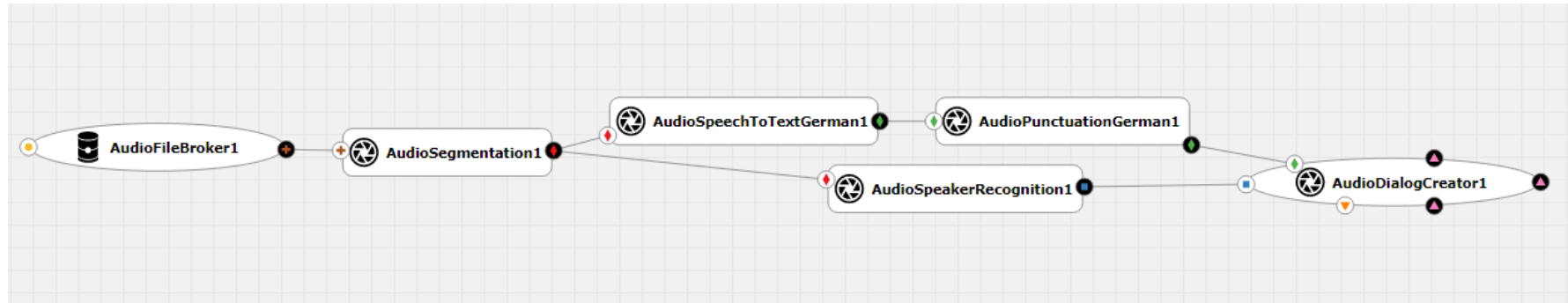
- Audio Pipelines composed of re-usable building blocks
- Visual Editor checks for compatible interfaces

AI4EU Container Specification
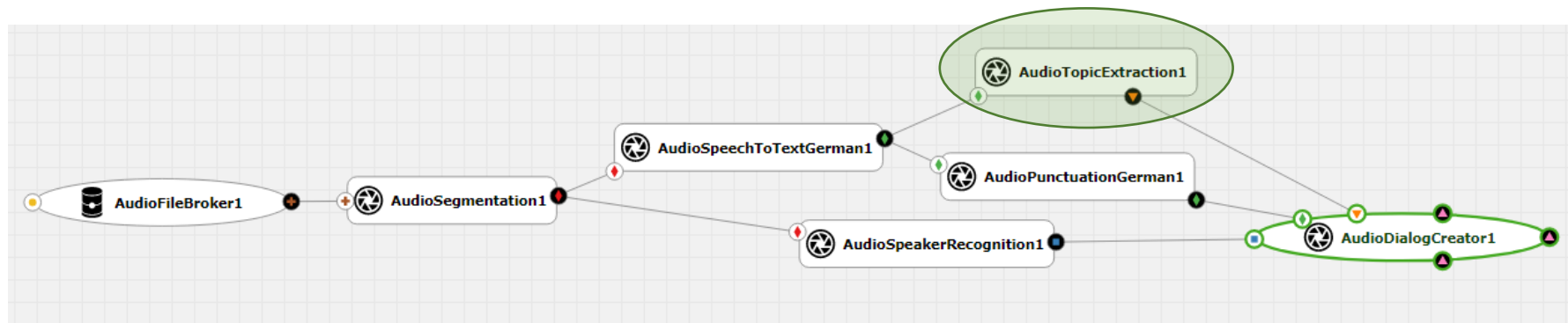
- Docker container

- Protobuf specification of public interface

- gRPC communication

- optional Web-UI for human interaction

- Based on free / open source technologies

- Recommendations for scalability, training and GPU-Support

- Support for gRPC streaming

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


// 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);
}
```

No Lock In: It's Docker

- AI4EU Experiments containers are standard Docker containers with additional properties
- They can be run and used completely independent from AI4EU Experiments in common environments:
  - Kubernetes
  - Docker-Compose
  - Standalone Docker

Intellectual Property

Open Source and commercial content is supported

- Each published tool must have a license
  - Open Source or
  - Commercial / Proprietary are welcome
- AI4EU Experiments stores only metadata
  - references to docker images
  - descriptions / documentation
  - no binaries

AI4EU Experiments Architecture

Central AI4EU Experiments Instance:
Model Catalog and Visual Composition

IT Domain of User:
Multiple Kubernetes Execution Environments

UI

Core Services

Export

Execution

Portal

Onboarding

Federation

Parallel
Orchestrator

AcuCompose

Commom Data
Service (CDS)

Kubernetes
Deployer

Pipeline Node

Deployment

MLOps

Cluster can be
private with access
to confidential data

Can easily
integrated with
MLOps workflows

Modular and clean microservices system

Databases

MariaDB

Nexus

Minikube
Kubernetes
Kubernetes+GPU

Datasets and docker images are only referenced and not stored inside the platform

8

Unmatched Scalability

Following the AI4EU Experiments Container Specification based on Docker, **the same Docker Image** can really scale out:

- **Small**: deploy on Minikube on a Laptop or PC

- **Standard**: deploy on medium Kubernetes clusters for standard use cases

- **Big**: deploy on a big Kubernetes cluster with many CPU-Cores and memory for massive concurrency inside the container as well as running many containers in parallel

- **Acclerate**: modern tools like PyTorch or Tensorflow can detect a GPU and use it if available

- **Model Training:** if the model exposes training service

Benefits of gRPC Communication

Using gRPC for the communication between the AI models has some important advantages:

- programming language independet (use the language you prefer)

- use your preferred tools and libraries inside the container

- simple

- standardized

- fast and efficient

- Communication stubs can be created dynamically at run time which enables generic orchestration/execution of pipelines

Advanced Orchestrator

- The orchestrator dispatches messages according to the pipeline topology

- Thanks to gRPC/Protobuf the communication stubs can be created on the fly

- No need for manual "plumbing" / coding

- Streaming support for event based use cases

- Certain cyclic topologies

- Parallel execution of messages

- Tool and programming-language agnostic

- **No other platform offers a similar feature**
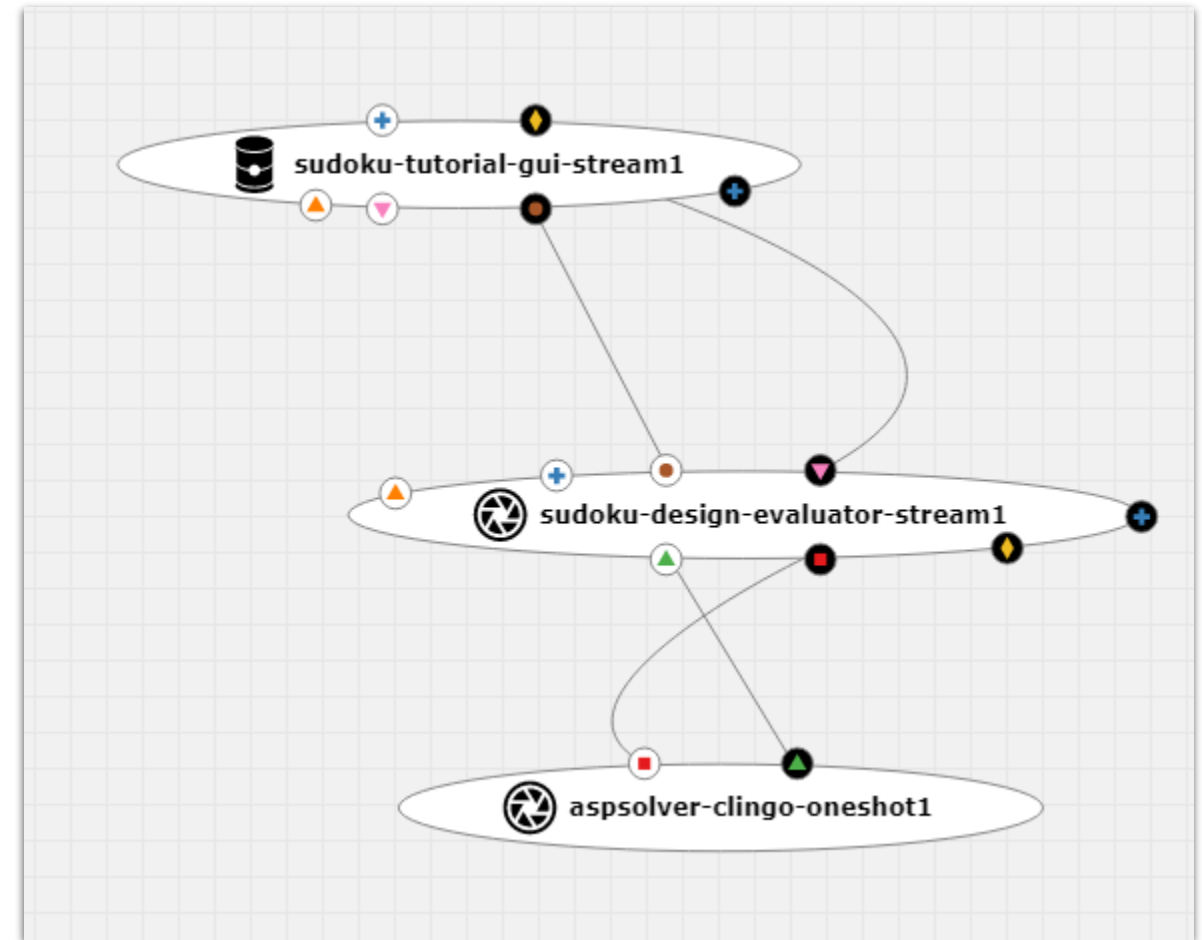
Training Pipelines
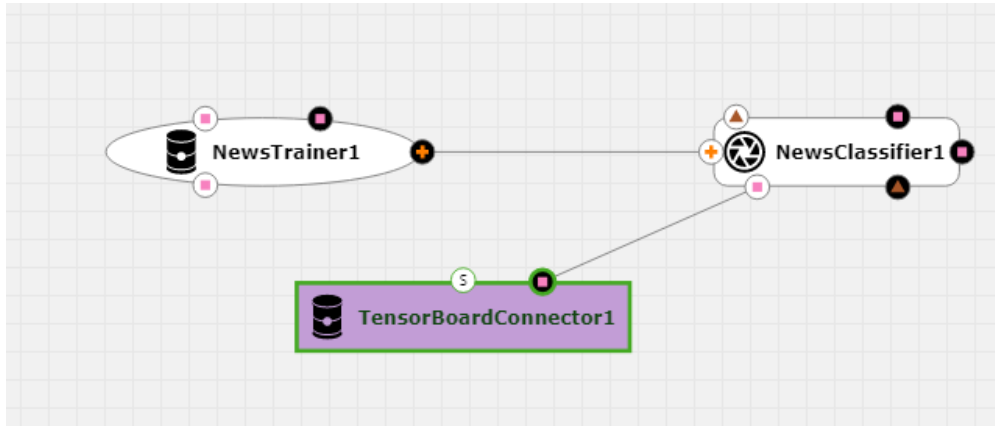
- Training controlled by dedicated Nodes
- Model node needs to provide training service
- Same docker container can be used for training and application
- Separate containers for application and training also possible

```
// similar to keras model.fit
message TrainingConfig {
  string training_data_filename = 1;   // .npz file
  string training_labels_filename = 2; // .npz file
  int32 epochs = 3;
  int32 batch_size = 4;
  double validation_ratio = 5;
  string model_filename = 6;
}


service Classifier {
  rpc classify(NewsText) returns(ClassifiedData);
  rpc trainModel(TrainingConfig)  returns(TrainingStatus);
  rpc setConfig(ConfigResponse)  returns(ConfigResponse);
}
```
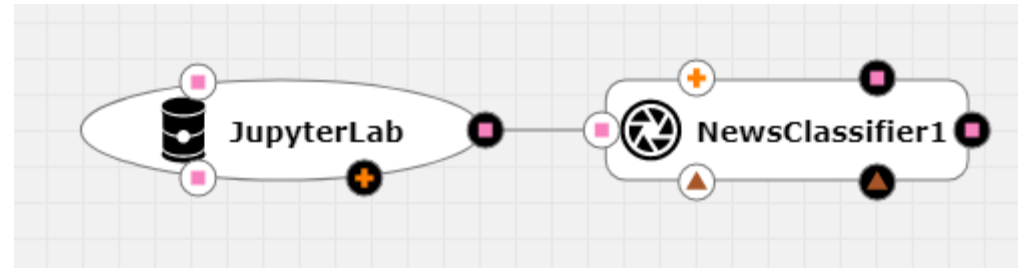
Utility Nodes
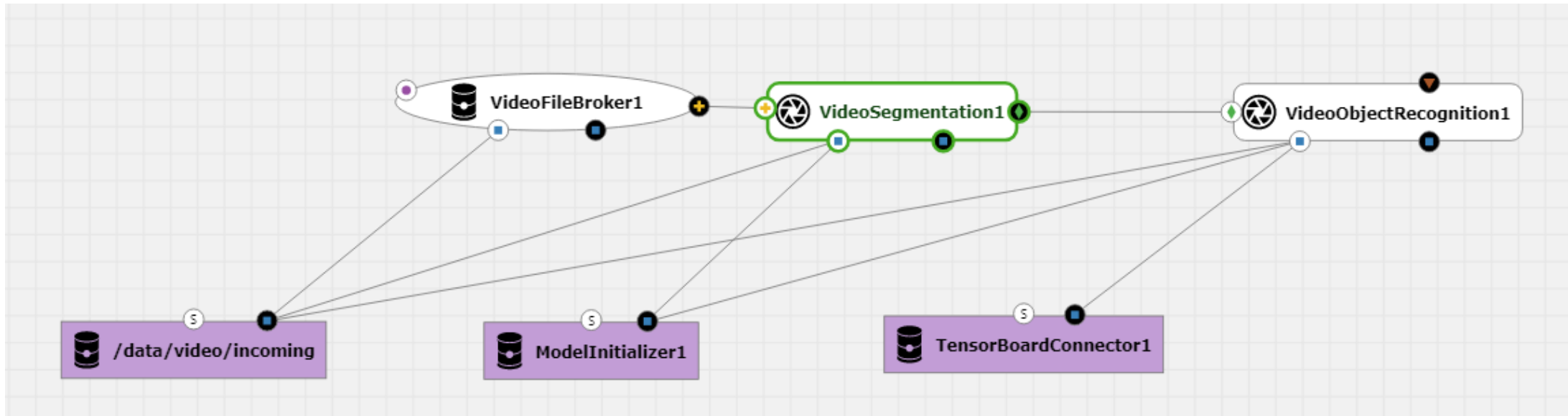
- Shared folders
- Model Initialization (planned)
- Diagnostic Tools (e.g. Tensorboard)

Jupyter Connect

- Work interactively with a model in a Jupyter Notebook

- The Jupyter node can be automatically generated thanks to protobuf and gRPC tools

- "One Click" deployment

- In the notebook, the connection to the model is already established

- Explore model properties with standard python tools like matplotlib, pandas or numpy

- planned feature

Research Use Cases
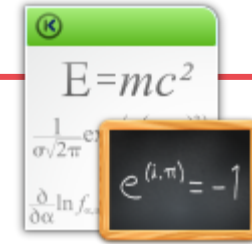
- Quickly configure and deploy training pipelines

- Hybrid AI Pipelines that combine symbolic and sub-symbolic tools

- Reproducability of pipeline runs (planned feature)

- Reusable building blocks for explainable or verifiable AI

- AutoML: reconfigure the topology and parameters of a pipeline for each iteration (planned feature)

Educational Use Cases

- Explain AI concepts using pipelines
- Exchange solutions
- Hands on excersises to implement parts of a pipeline
- Live classes and excercises with Jupyter Connect
- Collaborate on solutions

Playground and Demo Space

Thanks to the collaboration with KI.NRW there will be a

- Playground
- Demo Space
- for Models and Pipelines
- Also available beyond the project

Benefits for AI4Media Partners

Going beyond state of the art with the help of AI4EU Experiments

- Hybrid-AI components
  - define interfaces to combine symbolic and subsymbolic models
  - proofs of concept for interaction of DL and knowledge
  - add reasoning or SAT-Solvers to DL solutions
- Auto-ML
  - automate hyperparameters of a whole pipeline
  - define ways to Hybrid-Learning
- Cognitive Architectures
  - find the building blocks
  - specify common interfaces

Evolution of our platform

- AI4EU Experiments should be the beginning of an evolution

- Contribute and adopt it to your needs

- Exchange ideas and concepts as runnable, reusable artifacts

Useful Links

- AI4EU Experiments: https://aiexp.ai4europe.eu/#/home

- AI4EU Experiments Tutorials: https://github.com/ai4eu/tutorials
    - https://www.youtube.com/playlist?list=PLL80pOdPsmF6s6P6i2vZNoJ2G0cccwTPa

- AI4EU Homepage: https://www.ai4europe.eu/
    - Catalog: https://www.ai4europe.eu/research/ai-catalog
    - News & Events: https://www.ai4europe.eu/news-and-events

Most Important, the container specification:

https://github.com/ai4eu/tutorials/tree/master/Container_Specification