Al-based assistants in PUNCH2

Victoria&Arman, March 2025

Motivation

- The infrastructure is becoming too complex
- Variety of workflows: data analysis and simulations, reproducibility and provenance, [meta]data management, reporting, support...
- Different fields with similar problems
- Each field of research and project role has its own "Knowledge Domain":
 - PUNCH Communities: astrophysics, APP, PP, NP
 - PUNCH roles: researcher (beginners and advanced), data stewards/engineers, administrators, managers...
- Assistance is needed

Proposal

Vision: AI-assistant for PUNCH4NFDI SDP — chat bots based on custom knowledge can improve daily productivity.

- 1. AI based workflows and DRP generation
- 2. Al based code optimisation

1. Al based workflows and DRP generation:

In each community we build "Knowledge Bases":

- Astroparticle data analysis workflows: simulations (?), analysis of open datasets and plotting...
- Astronomy: join different tables from different data sources and generate plots
- Particle physics (DESY...): get data->run algorithms->produce DRPs
- Nuclear use case ...

=>All of them require special knowledge of experts

The goal is to build Based on "knowledge of experts" and "Knowledge Bases" and use them as a source for AI Assistants.

More for reference: proof of concept on generating REANA workflows using natural language

BS Thesis by ETL: <u>https://github.com/etlstrauss/bachelor-thesis-public</u>

2. Al based code optimisation:

- Optimize the current codes for performance and assist users for changes in order to fit within the PUNCH2 compute4punch and storage4punch infrastructure.
- This type of assistant may improve the power consumption or enhance productivity.

AI Assistants can be tightly integrated within the User Interfaces and Federated infrastructure.

Use Case: Astroparticle physics - AI Assistant for KASCADE Data Analysis

1. Current state

Context:

- <u>KCDC</u> provides open-access astroparticle physics data by <u>KASCADE experiment</u>
- Data users include professional and early career astroparticle physicists, students, and machine learning researchers
- Complex workflows involve simulation, data retrieval, processing, transformations and visualization

Challenges:

- Domain expertise is required to efficiently analyze the data
- Manual workflow setup (e.g., simulations, event selection, analysis) is time-consuming
- Accessing and using data correctly might require assistance for beginners

Use Case: Astroparticle physics - AI Assistant for KASCADE Data Analysis

2. Our solution

With AI-based Assistant:

- Natural Language Queries: Users can ask, "Retrieve high-energy air showers with iron primaries," and receive pre-filtered datasets
- Workflow Generation: AI helps construct REANA workflows for reproducible cosmic-ray simulations and data analysis
- Automated Data Processing: AI suggests optimal analysis pipelines based on expert knowledge from PUNCH4NFDI Knowledge Base
- Code Optimization: AI assists with scripts refactoring to run code efficiently on PUNCH2 infrastructure, supporting green computing

Benefits:

- Lower entry barrier for new users and interdisciplinary research
- Improved support for each group of users
- Higher productivity through faster, AI-guided analysis for experts
- Enhanced FAIR data reuse within PUNCH4NFDI

Technology

We are going to use modern LLM technologies like: RAG and Fine Tuning of LLMs for creating, managing and publishing our Knowledge Bases.

Data protection: In order to follow the data protection and security guidelines we are going to use OpenWeight models publicly available for self hosting. The software stack will be developed in modular way, so it can be hosted not only on the PUNCH infrastructure.

More details in the discussion document:

https://docs.google.com/document/d/1JQq8e-ImGYnW5o97Q383sOyiYJU5SYZsb0X-f1qROuk/edit?us p=sharing

LLM assistant

reana				Tutorial Abo	ut Contact Imprint 🔆
	HW Hello World! This example prints a simple 'Hello World' message through REANA.	Control Relation	Hello World!! This example prints a simple 'Hello World' message through REAVA.	Coom in Reaso Notebook This example shows how to get data from a remote database, in this case the AP StarHorse data Read more	
	0.8	≡ 0	0	8 0	
	Contribution of the second sec	Cipen in Reave HW Hello World!!Hello This example prints a simple Yello World message through REAVA. Mauris malesuada ante uma, a Read more	Costi in Receiv Upload & Downloa This example shows how to upload all files needed for the analysis (e.g. a catalog) and downloa Read more	Com It Read Access remote data This example show how to get data from a remote database, in this case the AP Starkforce data Read more	
	* 8	0 8	0 ¢ 8	5 Q	
	Con it Reads Retrieving astrono This example shows how access the data via TAPVO from external database and to use REANA to pl Read more	Corn in Resource	+ Add an example		
	٥	0 ≣ ¢			
localhost:3000/example/8					

Recreating or Analysing



Use Case: Al Assistant for Gaia Data Analysis

- Gaia provides high-precision astrometric, photometric, and spectroscopic data for over a billion stars.
- Upcoming GAIA DR4 data >1PB data
- Handling **large datasets** in Observational Astronomy efficiently on distributed infrastructure is non-trivial.
- Complex queries using IVOA protocols require domain expertise.
- Users need to **join multiple catalogs** (e.g., Gaia + SDSS) for cross-matching.

Solution: Al-based Assistant for Gaia & IVOA

- **Natural Language Queries**: "Find all Gaia DR3 stars with parallax > 10 mas and cross-match with SDSS spectroscopic data."
- Automated Query Generation: Al translates requests into TAP/ADQL queries for Gaia archives.
- Workflow Assistance: Al helps construct REANA workflows for cross-matching and visualization.
- **Code Optimization**: Al assists in restructuring ADQL queries and optimizing data handling on PUNCH2 infrastructure.
- Impact:

Easier access to Gaia and other astrophysical data for astronomers and non-experts.

Faster, AI-guided dataset exploration.

Seamless integration of IVOA standards in astrophysical research.

Efficient computation using PUNCH4NFDI infrastructure.

Requirements

 $2\,\text{FTE}$

1-researcher(domain expert)

1-developer

