AI developments and platforms for CRG, industrial, and medical applications
KT, synergies, and spin offs

Diogo Reis Santos

09.09.2024

Acknowledgments

Data and Image Analysis Team (DIAG)

TE CRG

Luigi Serio

TE CRG ML

Lorenzo Gusti

Albert Aillet

IT

Heloisa Barbosa

Paolo Cacace

Andrea Protani

KT

Paolo Barba

SUMMARY

all externally funded and covered by KT agreements approved by the ATS sector and TE department

- **CRG**
  - Anomaly Detection and Maintenance Optimization in Large-Scale Cryogenic Compressor Systems
  - Modeling the LHC's BHX using graph neural networks
  - PenguinGPT, a chatbot for Cryogenic applications

- **CAFEIN - Federated network platform for developing and deploying AI-based analysis and prediction models**

- **Medical Applications**
  - Risk stratification for breast and prostate cancer based on WHO IARC's EPIC data
  - TRUSTroke and Umbrella, a holistic approach to prevention, treatment, and management of stroke
  - WHO Market Analysis Platform (MIP) for Prevention and Mitigation of Supply Chain Disruptions
  - STELLA, Smart Technologies to Extend Lives with Linear Accelerators

Anomaly Detection and Maintenance Optimization in Large-Scale Cryogenic Compressor Systems

- **Problem formulation**
  - Helium Compressors occasionally fail before EUL
  - An external company evaluates compressor vibration ~1/month
  - Evaluate current state (not future)
 - Occasionally, compressors fail in <~ 1 month
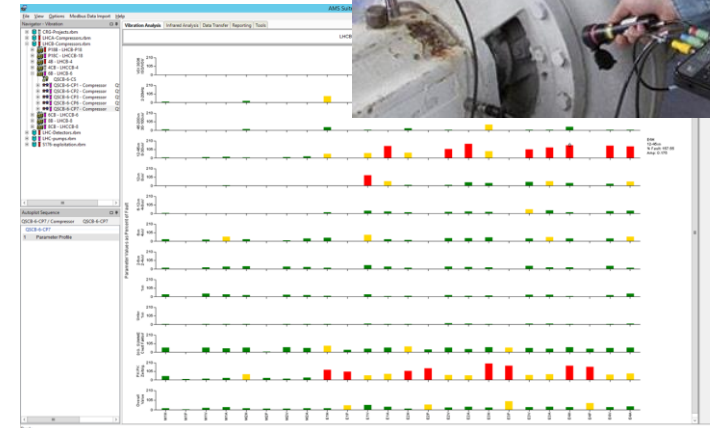
- **Aim**
  - Improved predictive system based on AI
  - Predictive maintenance and RUL (Remain Useful Life)
  - Capable of detecting issues up to 1 year in advance
  - Integrated into a future online system

Anomaly Detection and Maintenance Optimization in Large-Scale Cryogenic Compressor Systems
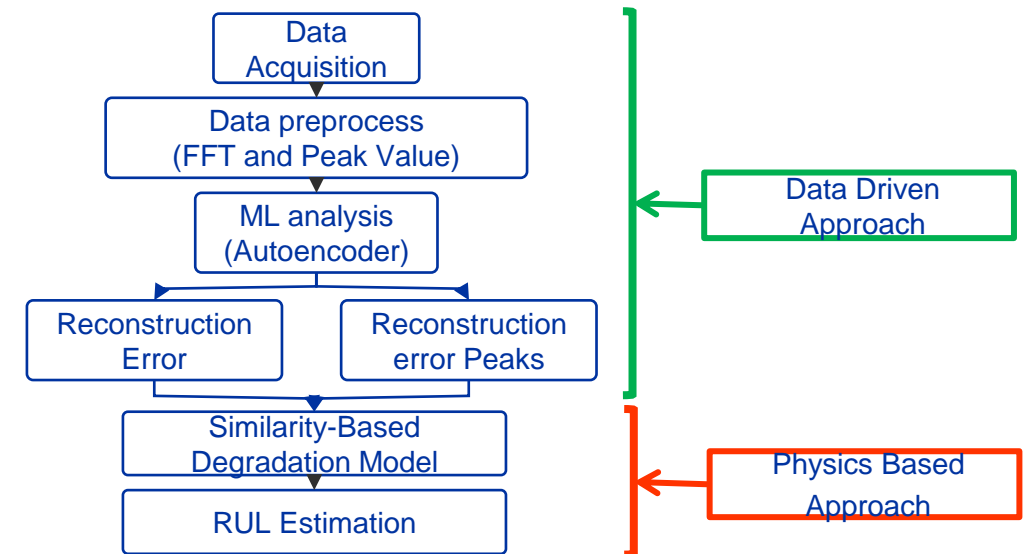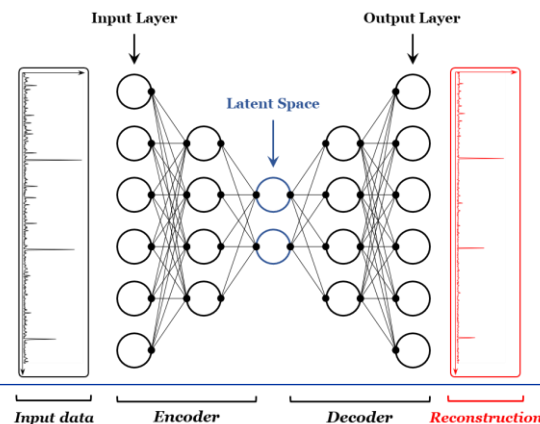
- **Data preparation**
  - System/position-based to Asset-based data
  - Labeling data based on real historical data
  - Reverse engineering and extracting vibration data
  - Merge with PyTimber data
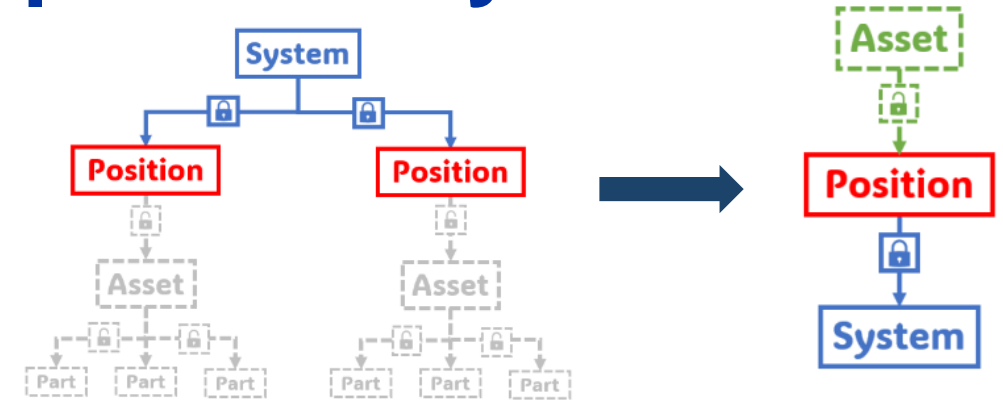
- **Model**
  - Autoencoder-like model trained on normal operational data
  - State model for prediction
  - Trained on data 2016-2020
  - Evaluated on 2020+

Anomaly Detection and Maintenance Optimization in Large-Scale Cryogenic Compressor Systems

Results

- Good performance for "Normal" operation
  - F1-score 95% to 96% for type B and type H

- Improved performance for "Warning"
  - F1-score 55% to 67% for type B
  - F1-score 57% to 60% for type H

- Addition of "Critical" alarm
  - F1-score 0% to 100% for type B
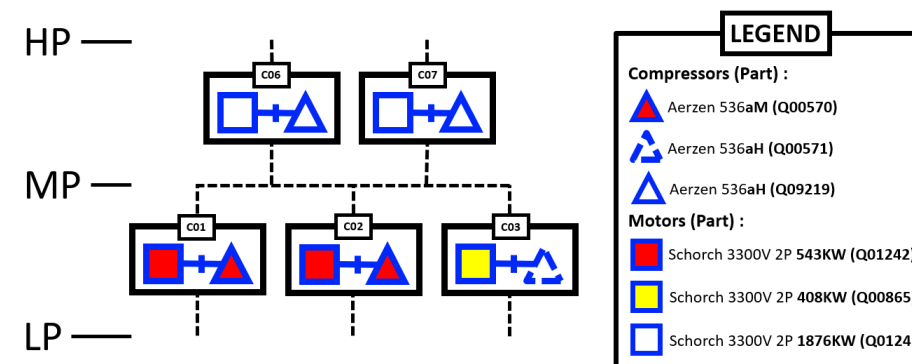 - F1-score of 73% to 86% for type H

Challenges

- ETL pipelines
- Adaptation/embedding to/in online sensor (CAFEIN-IOT)
- Anomaly detection to other components (supervised /unsupervised)

RUL Based 3 Class Classification of Type-B Compressors (C01-C02)

Category	Precision		Recall		F1-score		Support
Model	Our Model	Expert Labels	Our Model	Expert Labels	Our Model	Expert Labels	
Normal	**0.95**	0.94	**0.98**	0.96	**0.96**	0.95	225
Warning	**0.77**	0.55	**0.59**	0.55	**0.67**	0.55	29
Critical	**1.00**	0.00	**1.00**	0.00	**1.00**	0.00	3
RUL-Model Output Summary				**Expert Classification Summary**			
Category	Precision	Recall	F1-score	Category	Precision	Recall	F1-score
Macro avg	0.91	0.85	0.88	Macro avg	0.50	0.50	0.50
Weighted avg	0.93	0.93	0.93	Weighted avg	0.89	0.9	0.89
Accuracy		**0.93**		Accuracy		0.90	

RUL Based 3 Class Classification of Type-H Compressors (C06-C07)

| Category	Precision		Recall		F1-score		Support
Model	Our Model	Exp-Labels	Our Model	Exp-Labels	Our Model	Exp-Labels	
Normal	0.95	**0.96**	**0.97**	0.93	**0.96**	0.95	183
Warning	**0.71**	0.50	0.52	**0.65**	**0.60**	0.57	23
Critical	0.75	**0.80**	**1.00**	0.67	**0.86**	0.73	6
RUL-Model Output Summary				**Expert Classification Summary**			
Category	Precision	Recall	F1-score	Category	Precision	Recall	F1-score
Macro avg	0.80	0.83	0.81	Macro avg	0.75	0.75	0.75
Weighted avg	0.92	0.93	0.92	Weighted avg	0.91	0.89	0.90
Accuracy		**0.93**		Accuracy		0.89	

Paolo Cacace et al, "Machine Learning Framework for Anomaly Detection and Maintenance Optimization in Large-Scale Cryogenic Systems, ICEC 2024

9 September 2024   6

Modeling the LHC's BHX using graph neural networks

- **Problem formulation**
  - The SHe cryogenic system is complex
  - Developing, maintaining, operating and improving it is not trivial

- **Aim**
  - Develop AI to assist the
    - Design, test, and validation
    - Advance control (JT valve)
    - Diagnosis

- **POCs**
  - Master thesis (BHX)
  - String CRG system
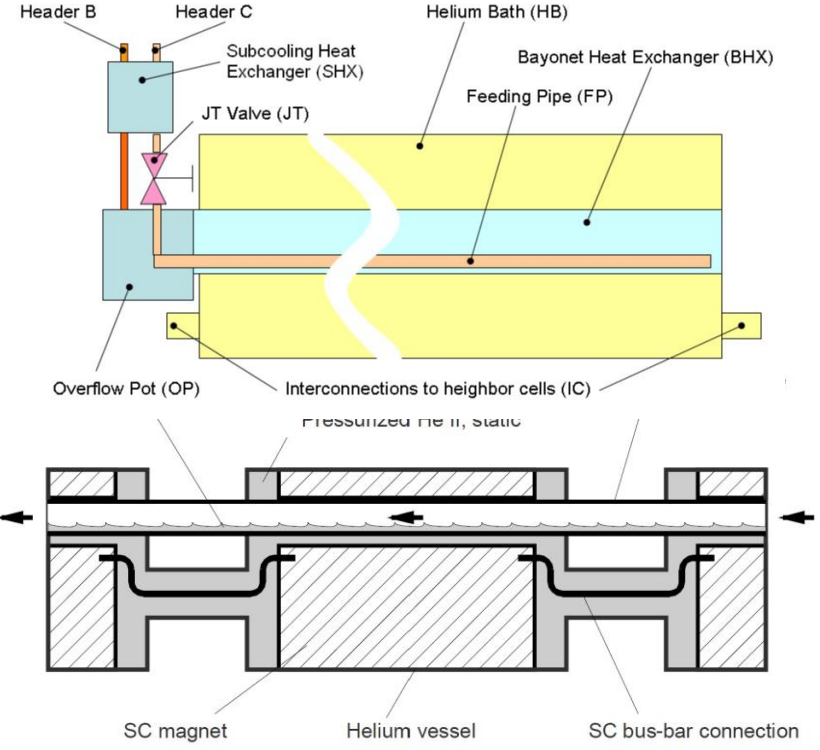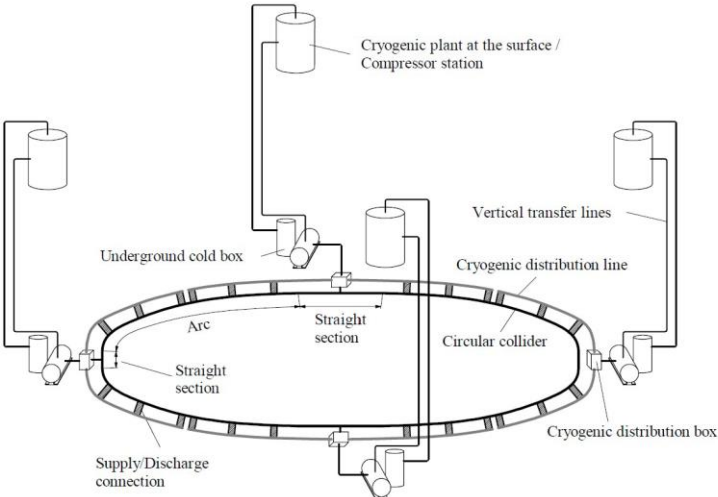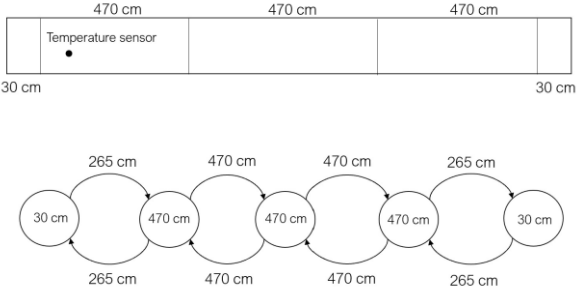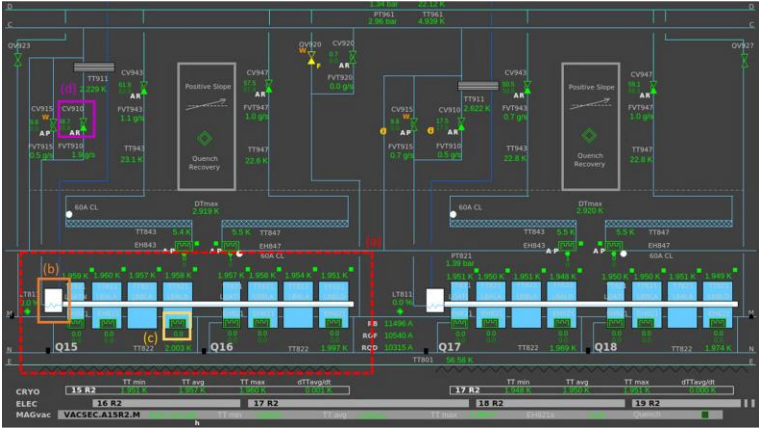
Modeling the LHC's BHX using graph neural networks

- **Current state**
  - SM18 LHC and HL-LHC string
  - Trained on data from 1.11.23 – 30.11.23 (PyTimber)
    - (cool-down)
  - Evaluated on data from 1.12.23 - 15.1.24 (PyTimber)
    - (warm-up)
  - Input -> SHe feeding valve position
  - Trained to predict temperature and pressures

- **Challenges**
  - Complete the GNN model of the BHX
  - Evaluation against SOTA (EcoSimPro)
  - Read from Cryogenic Diagram (schematic)
 - ETL from PyTimber
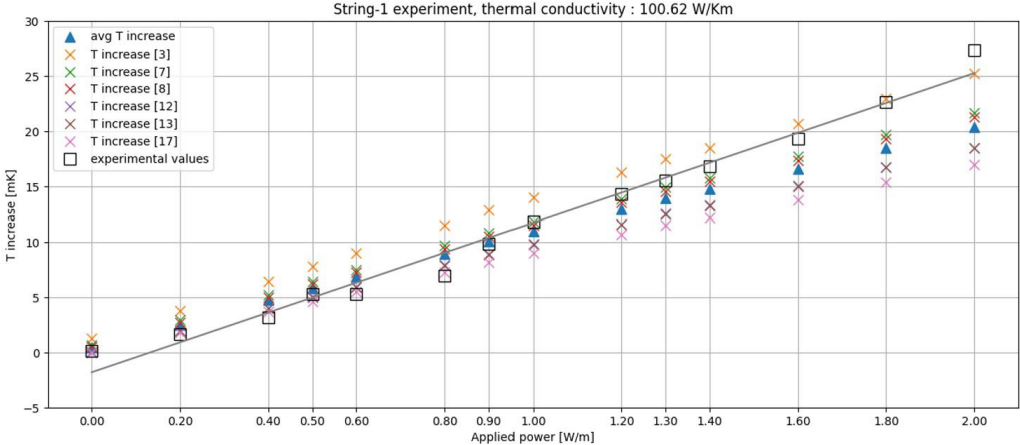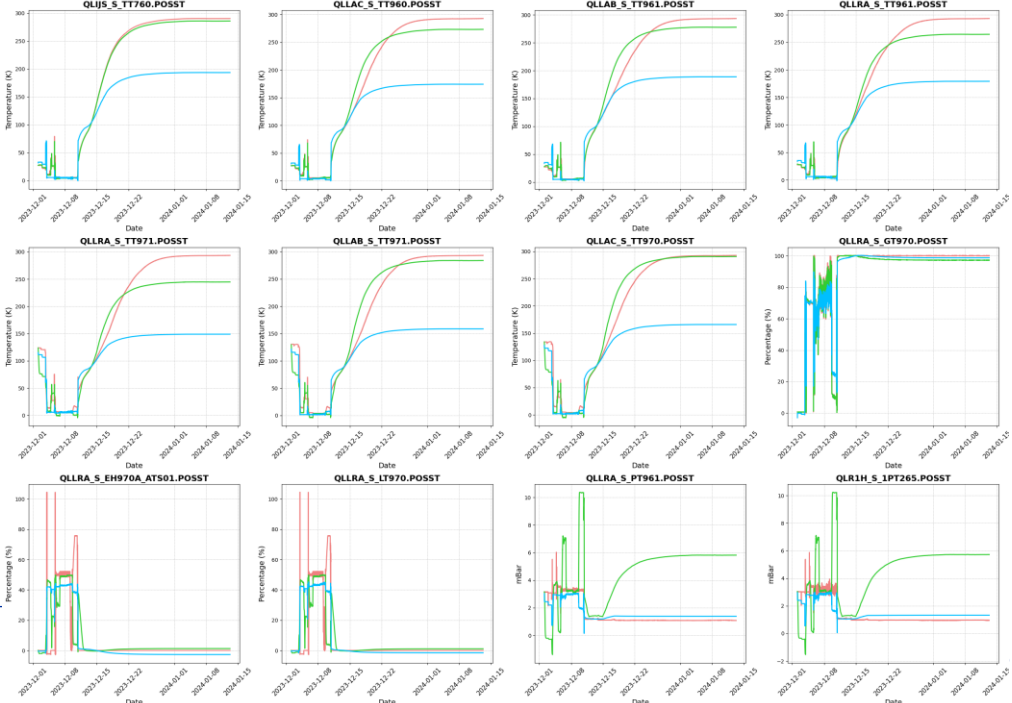
PenguinGPT, a chatbot for Cryogenic applications

- **Problem formulation**
  - Back-office users and field technicians need to perform elaborate queries to find information.
  - Information is spread across several systems and occasionally does not straightforward to find in a document.

- **Aim**
  - **NLP Integration**: Allows technicians to ask about position and asset statuses using conversational language
  - **CMMS Integration**: Connects to existing systems like EAM, EDMS, DFS, and CERNBOX.
  - **Key Capabilities**:
    - Query cryogenics equipment for positions, assets, and system details.
    - Retrieve data records, standard/custom fields, and associated materials (parts).
    - Access meter readings and interpret related work, such as task plans, PM schedules, and work orders.
    - View documentation linked to the EAM structure in EDMS.
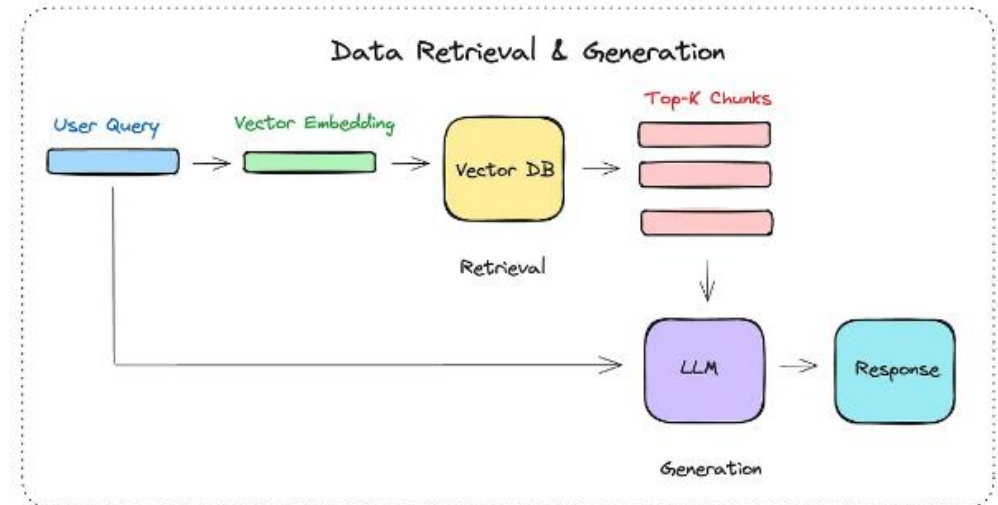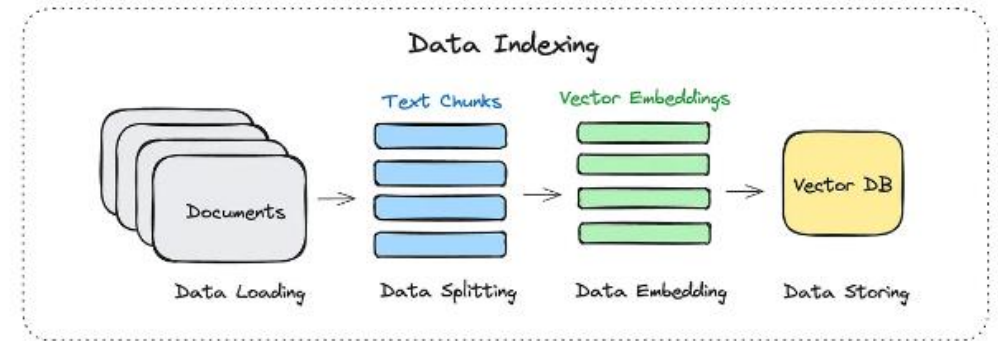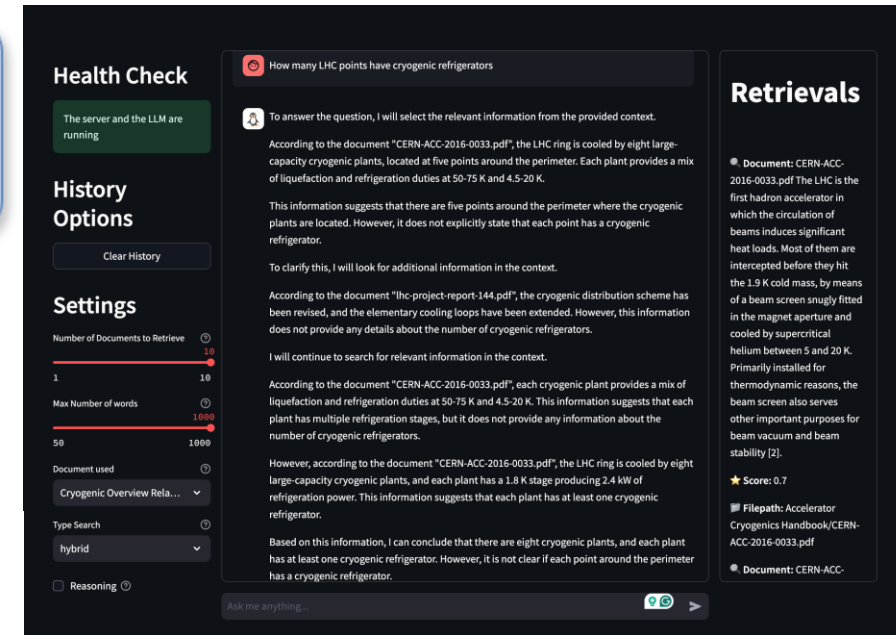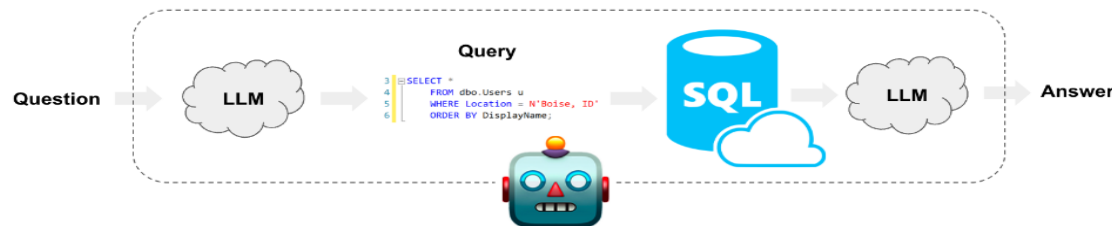
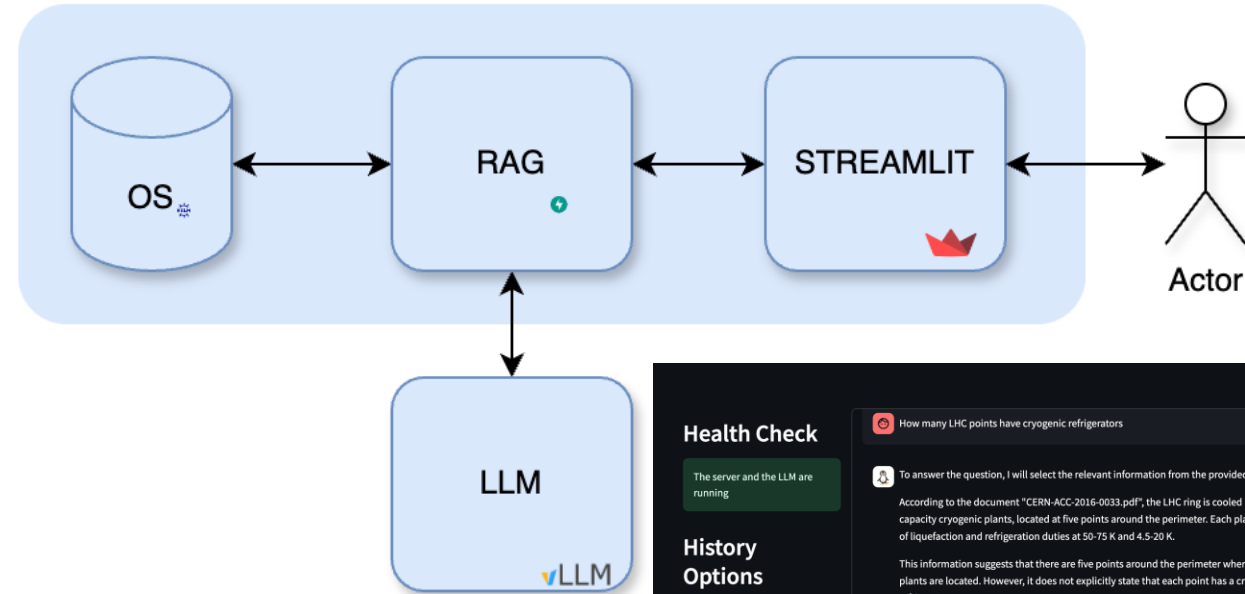# PenguinGPT, a chatbot for Cryogenic applications

- **Current State**
  - RAG system (for document search)
 - SQL
  - Summarization

- **Infrastructure**
  - OpenSearch
    - Document Store and Vector DB
  - LLM
    - Lamma-3.1-8B-Instruct
  - RAG
    - Developed by us (prompt engineering)
 - Streamlit
    - UI

PenguinGPT, a chatbot for Cryogenic applications

- **OpenSearch is great and available "as a service"**

- **LLM are challenging to deploy (GPU requirements)**
  - Could be shared and available "as a service"

- **Embedding models (sentence-transformers) have room for improvements**
  - CERN/CRG fine-tuned

- **LLM foundation models are evolving fast (3 models used in this project)**
  - CERN/CRG fine-tuned

- **Straightforward using Natural Language text**
  - Tables, images, diagrams, SQL still have room for improvement

- **Collaboration/integration with ML is crucial**
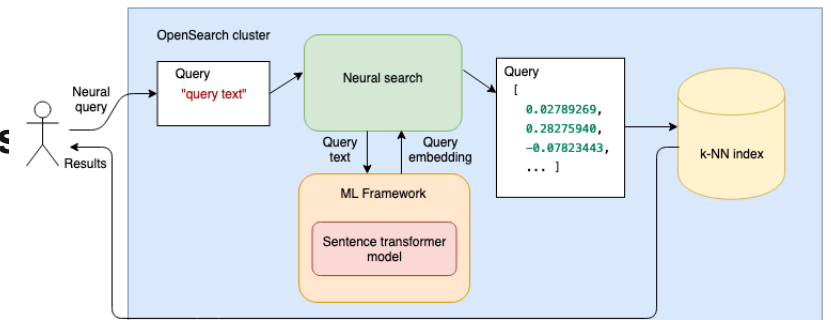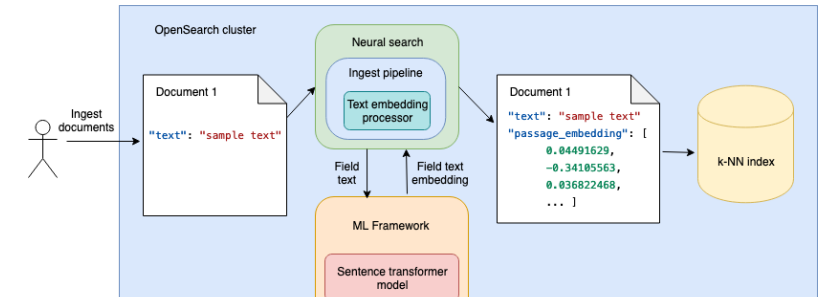
PRICE

USD per 1M Tokens; Lower is better
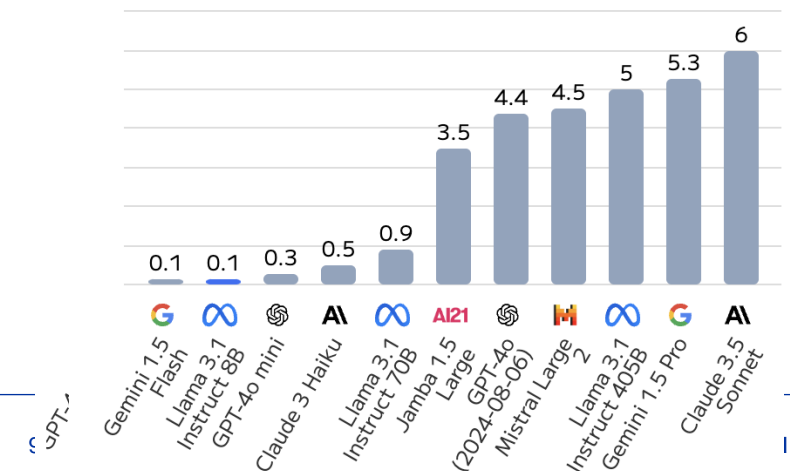
CAFEIN - Federated network platform for developing AI-based analysis and prediction

- **Computational Algorithms for Federated Environments: Integration and Networking**

- **Federated Learning (FL) platform developed at CERN**

- **Platform (Hardware and Software)**

- **Based on the MQTT protocol**

- **Design for production environments**

- **Maintained by us**
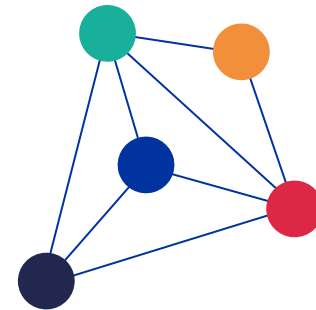
CERN
CAFEIN

CAFEIN - Federated network platform for developing AI-based analysis and prediction

- ## Screening tool brain MRI

  - *Decentralized Federated Learning for Healthcare Networks: A Case Study on Tumor Segmentation, B. Camajori Tedeschini, S. Savazzi, R. Stoklasa, L. Barbieri, I. Stathopoulos, M. Nicoli, L. Serio, January 2022, in IEEE Access, 10.1109/ACCESS.2017.DOI*

- ## EEG for stroke analysis

  - Andrea Protani et al ., Towards Explainable Graph Neural Networks for Neurological Evaluation on EEG Signals, HealthCom 24

- ## Atrial Fibrillation detection

  - Diogo R Santos et al ., Feasibility Analysis of Federated Neural Networks for Explainable Detection of Atrial Fibrillation, IEEE HealthCom 24
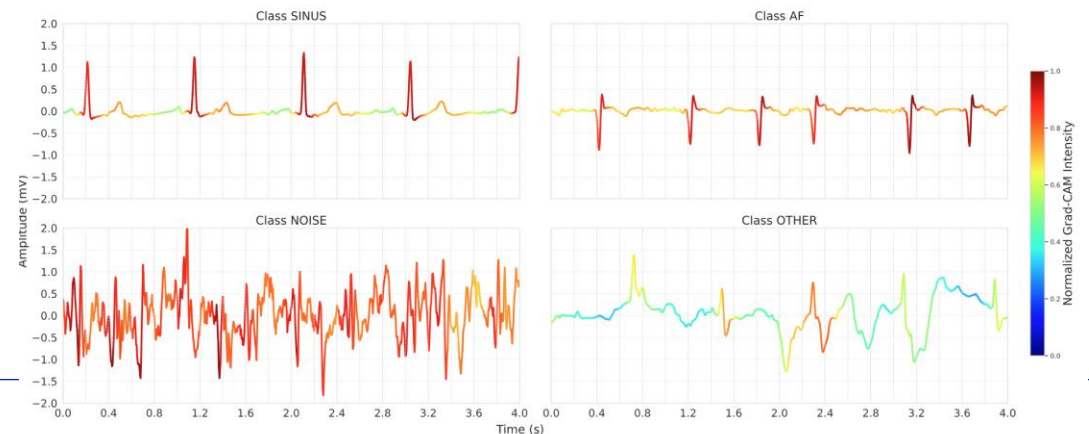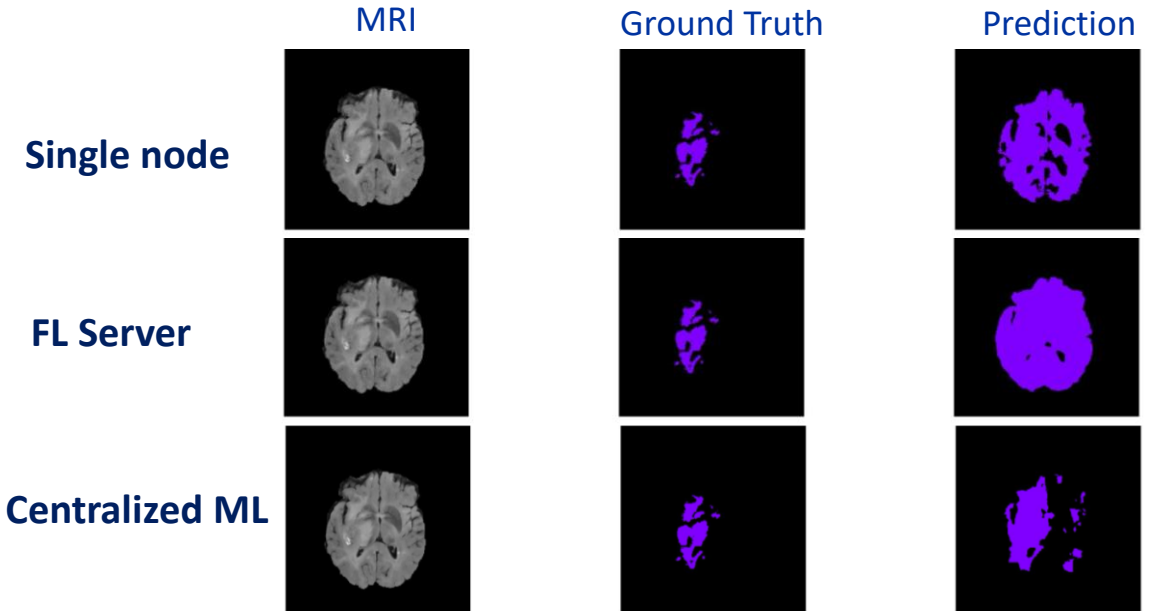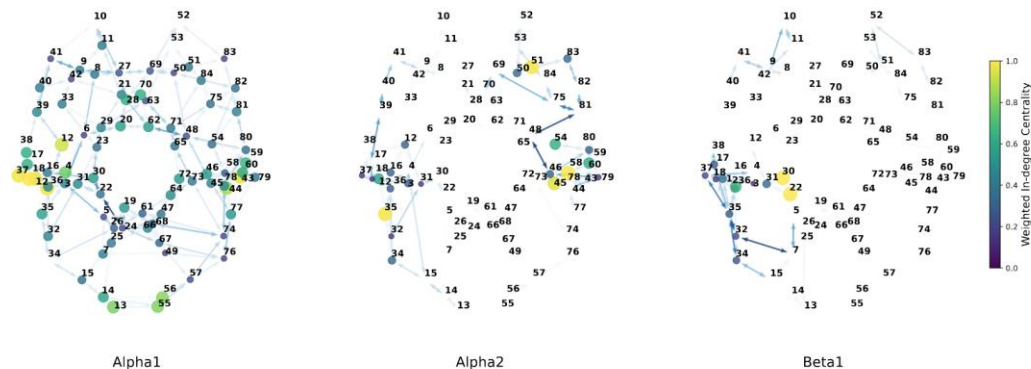
CAFEIN - Federated network platform for developing AI-based analysis and prediction

Initial R&D → Proof of Concept (POC) → Minimum Viable Product (MVP)

Scalability
- Multiple models
- Hundreds of clients
- Auto-scaling/load-balancing

Security
- Network security
- Model security

Usability
- User Experience (UX)
- Documentation

Reliability
- Fault Tolerance
- Backup and Recovery
- Testing

Performance
- Training times
- Model metrics
- Explainability
- Comparison to other platforms

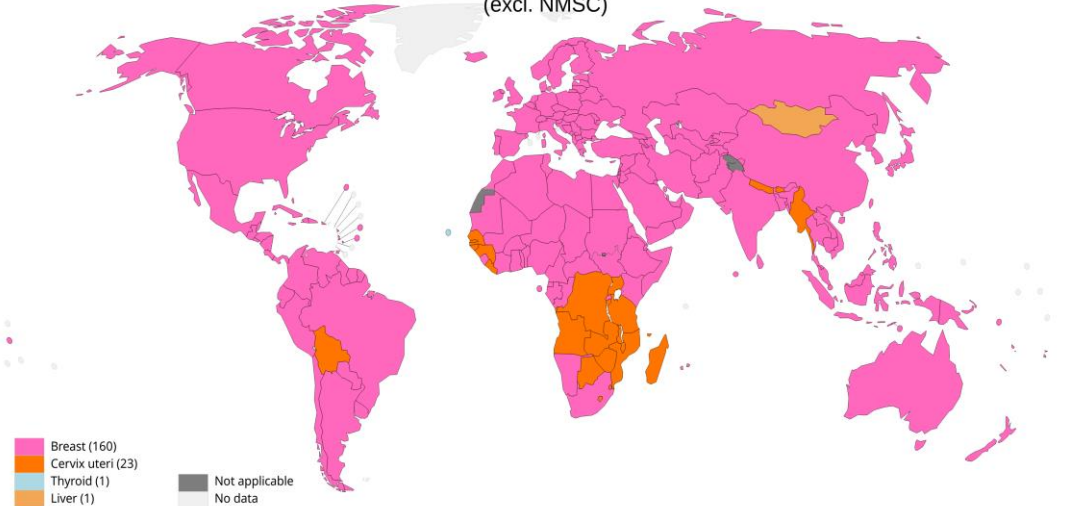# Risk stratification for breast cancer based on WHO IARC's EPIC data

- Most frequently diagnosed type of cancer in women.
- 2nd cause of cancer-related death among women.
- 2.3 million women were diagnosed / year
- 685.000 deaths /year

"Does screening work?"

shall move to

"For whom does screening work?"

Top cancer per country, estimated age-standardized incidence rates (World) in 2020, females, all ages (excl. NMSC)
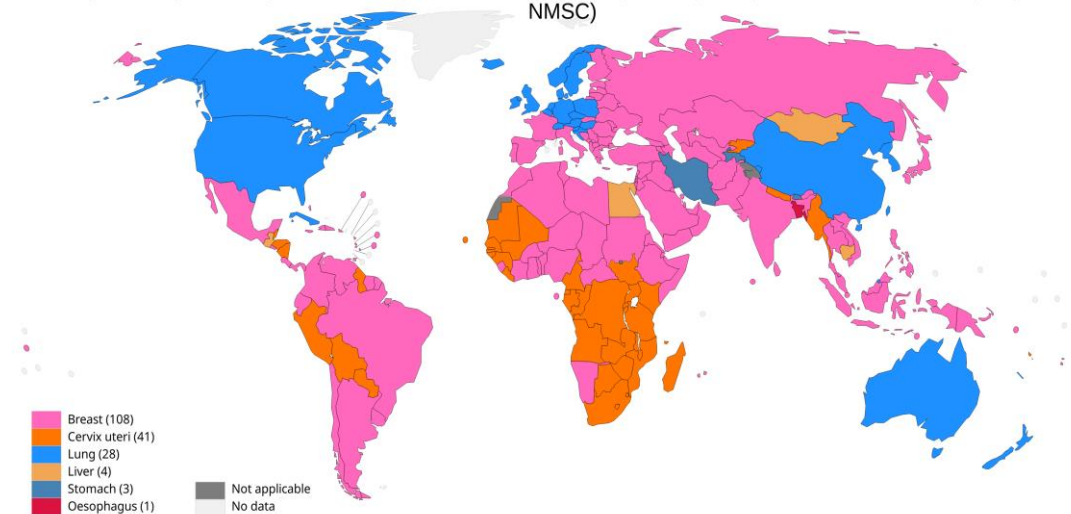
Breast (160)
Cervix uteri (23)
Thyroid (1)
Liver (1)
Not applicable
No data

Top cancer per country, estimated age-standardized mortality rates (World) in 2020, females, all ages (excl. NMSC)

Breast (108)
Cervix uteri (41)
Lung (28)
Liver (4)
Stomach (3)
Oesophagus (1)
Not applicable
No data

Risk stratification for breast cancer based on WHO IARC's EPIC data

- One of the most extensive cohort studies in the world
- 10 Western European countries
- More than 500.000 participants
- Follow up 20+ years
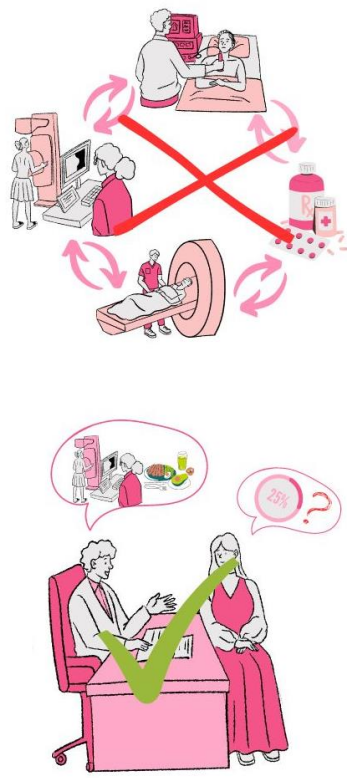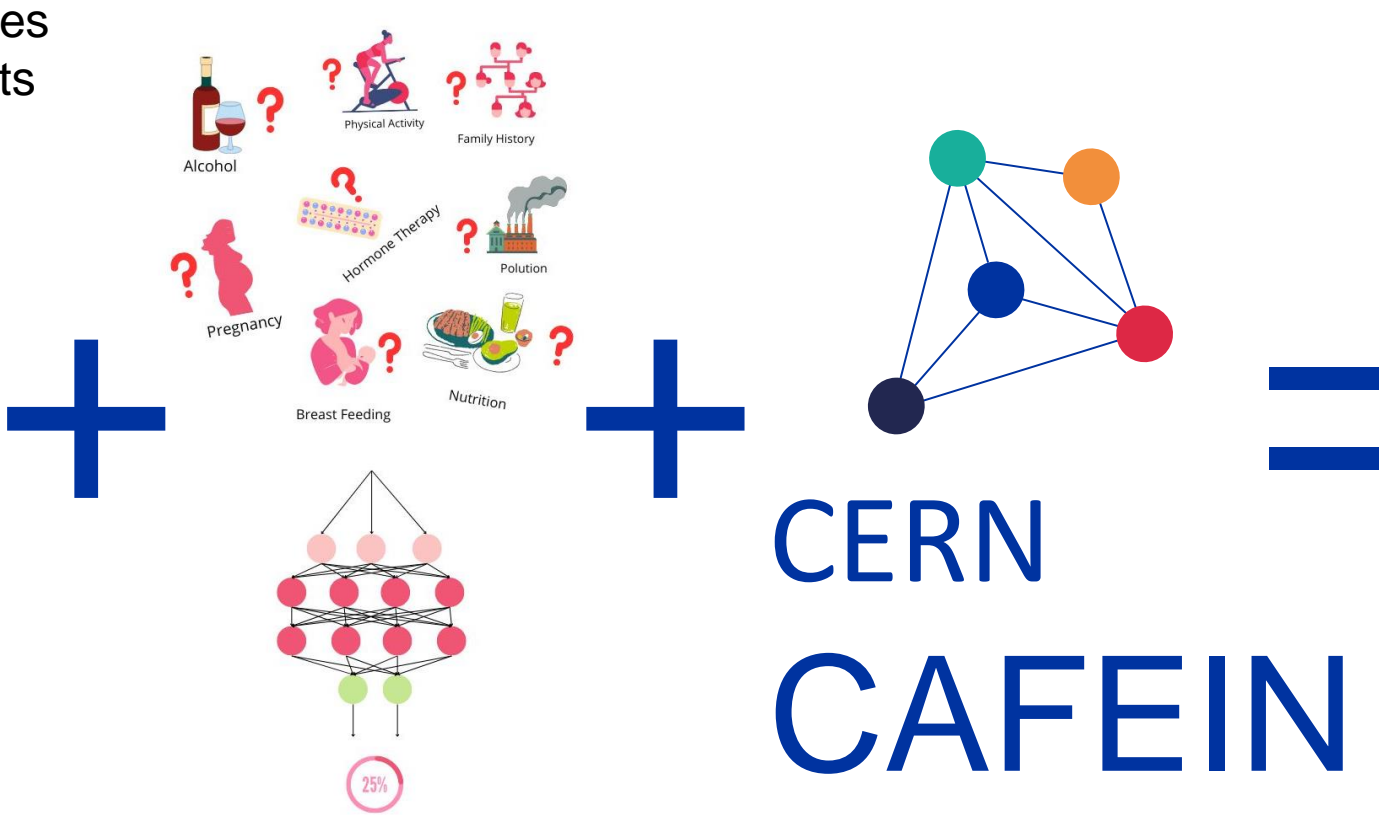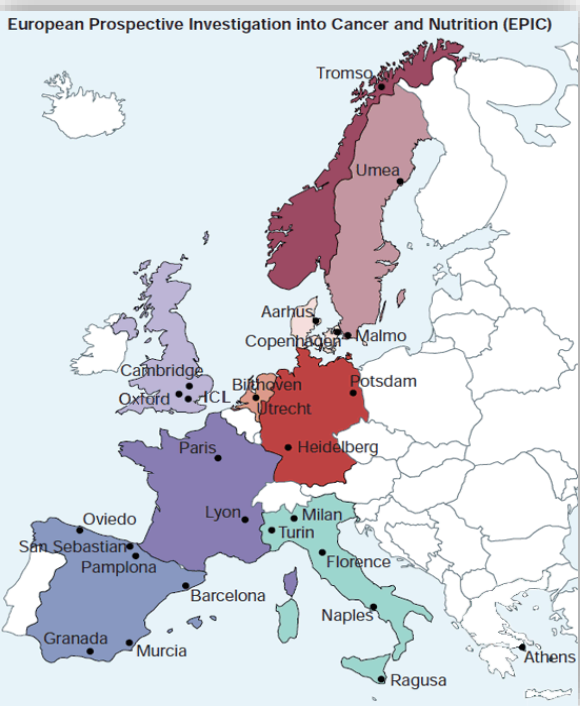
Risk stratification for breast cancer based on WHO IARC's EPIC data

TABLE II: Results for Gail, LightGBM Gail, and LightGBM All with Gail features for the test set.

Model	@ 0.5 Threshold			@ Max F1			PR-AUC	ROC-AUC	Brier
	Precision	*Recall*	*F1*	*Precision*	*Recall*	*F1*			
Gail	0.00	0.00	0.00	0.04	0.20	0.07	0.04	0.51	0.039
LGBM Gail	0.05	0.57	0.09	0.04	0.56	0.08	0.05	0.56	0.23
LGBM All	0.05	0.36	0.09	0.05	0.31	**0.09**	**0.06**	0.58	0.19

- **GAIL**
  - PR-AUC 4%

- **LightGBM**
  - PR-AUC 6% (+50%)

- **Explainability**
- Non-modifiable factors
  - Brain circulatory disease ↑
- Modifiable factors
  - Meat consumption (QgR0700) ↑ but very-high ↓
  - Alcohol, cigarettes/cigars ↑
  - Vegetables (QgE0201) ↓
  - Household physical activity ↓
  - Age of 1st pregnancy ↑
  - C/A Hormones (Menopause) ↑
  - Hip/Height measure ↑
  - Fat (fatty acid 22:5) (Docosapentaenoic acid) (intermediate fatty acid species between EPA and DHA) ↓
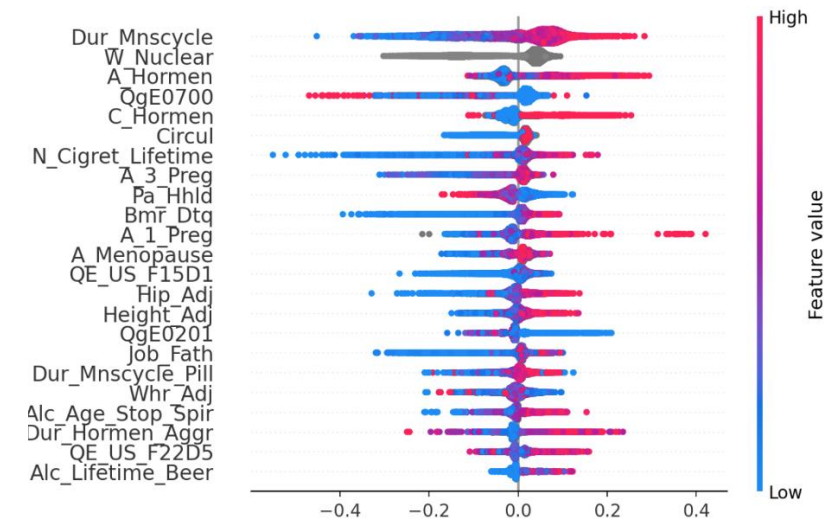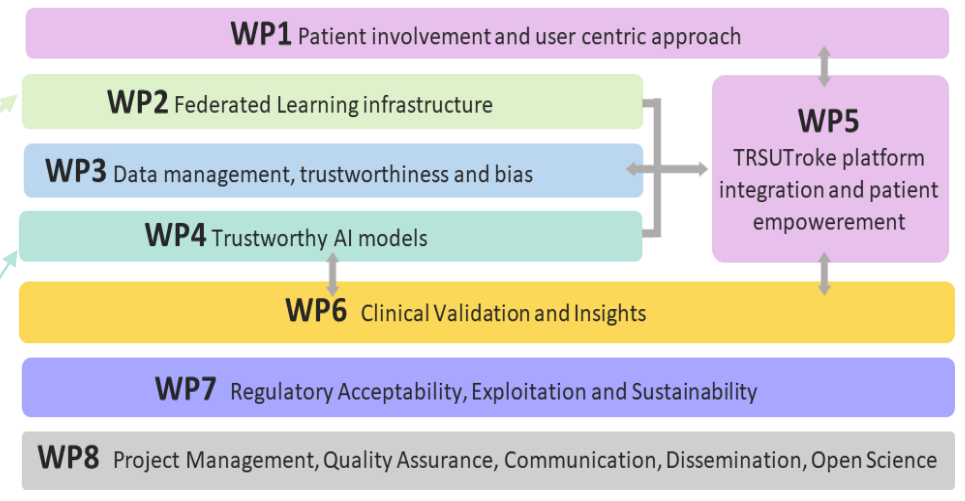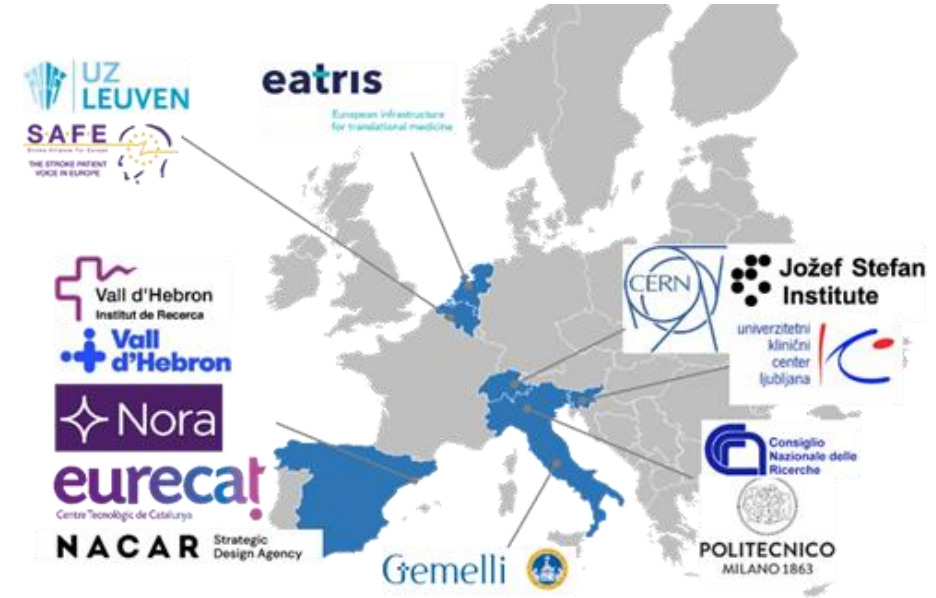

Fig. 3: Explainability of the oversampling LightGBM model for the test set.

TRUSTroke

- **Stroke is the leading cause of severe disability worldwide**
    - 1.1 m strokes/y in EU
    - 0.5 m deaths/y in EU
    - 9.5 m stroke survivors
    - 1st cause of disability

- **AI-tool based on the integration of clinical and patient-reported data**

- **Almost 10'000 enrolled patients' data will train algorithms over CERN federated learning platform**

WP2, led by CERN, is devoted to the design and development of the FL infrastructure, the implementation and validation of the federated system composed of different hospitals across Europe.

WP4, 5, 7 and 8, CERN participation

WP1 Patient involvement and user centric approach

WP2 Federated Learning infrastructure

WP3 Data management, trustworthiness and bias

WP4 Trustworthy AI models

WP5 TRSUTroke platform integration and patient empowerement

WP6 Clinical Validation and Insights

WP7 Regulatory Acceptability, Exploitation and Sustainability

WP8 Project Management, Quality Assurance, Communication, Dissemination, Open Science

TRUSTroke FL Infrastructure

Diogo R Santos et al ., "A Federated Learning Platform as a Service for Advancing Stroke Management in European Clinical Centers"

IEEE HealthCom 24

9 September 2024

20
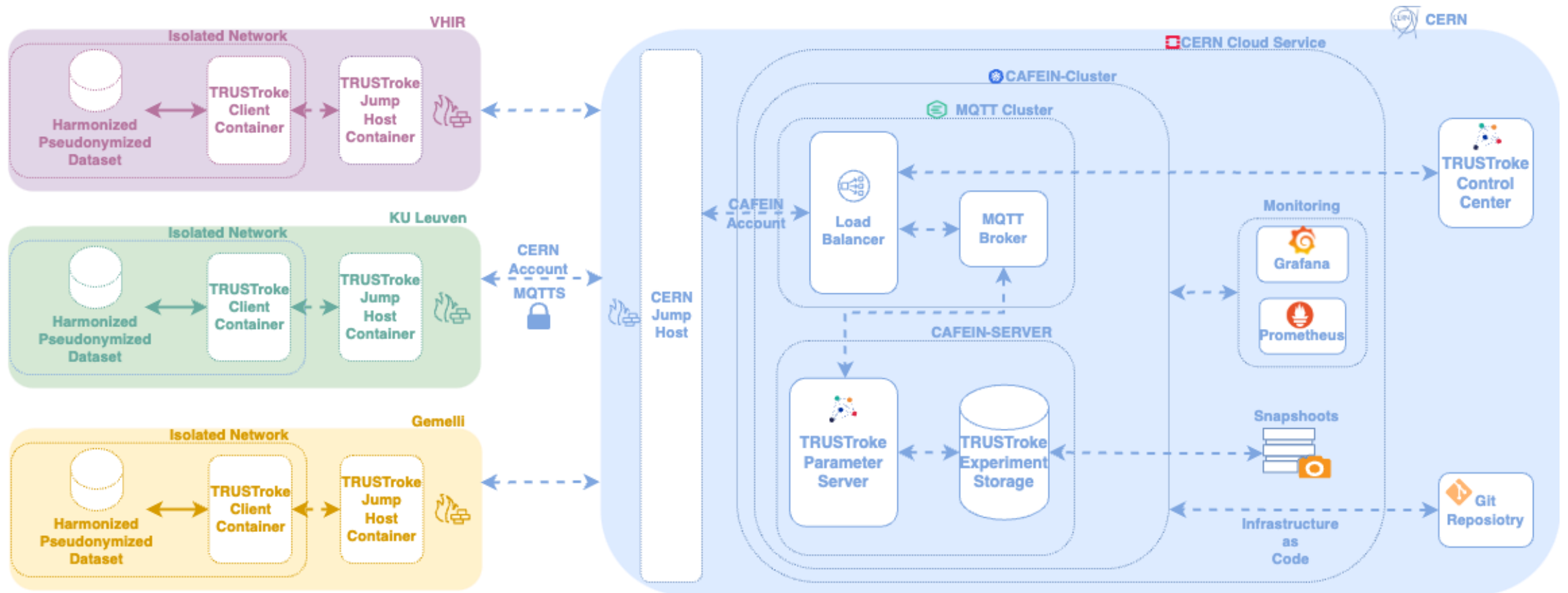
UMBRELLA

- A holistic approach to progress, reshape, and benchmark the overall stroke care pathway .

- Set new and improved standards of care in terms of primary and secondary prevention, rapid access to treatments, early accurate diagnosis, stratification, management and real-time monitoring, therapeutic targets identification, and rehabilitation, recurrent stroke, and related cardiovascular events.

- *TRUSTroke X 5*

- Oct 24 – Sep 29
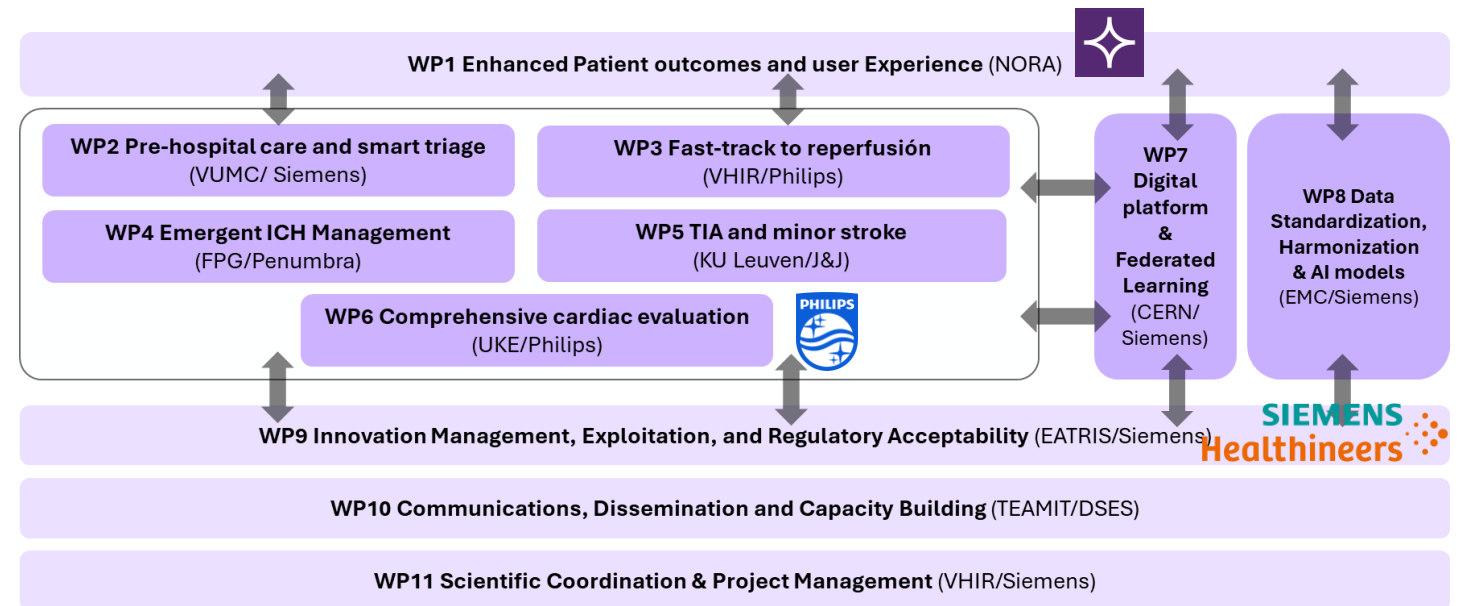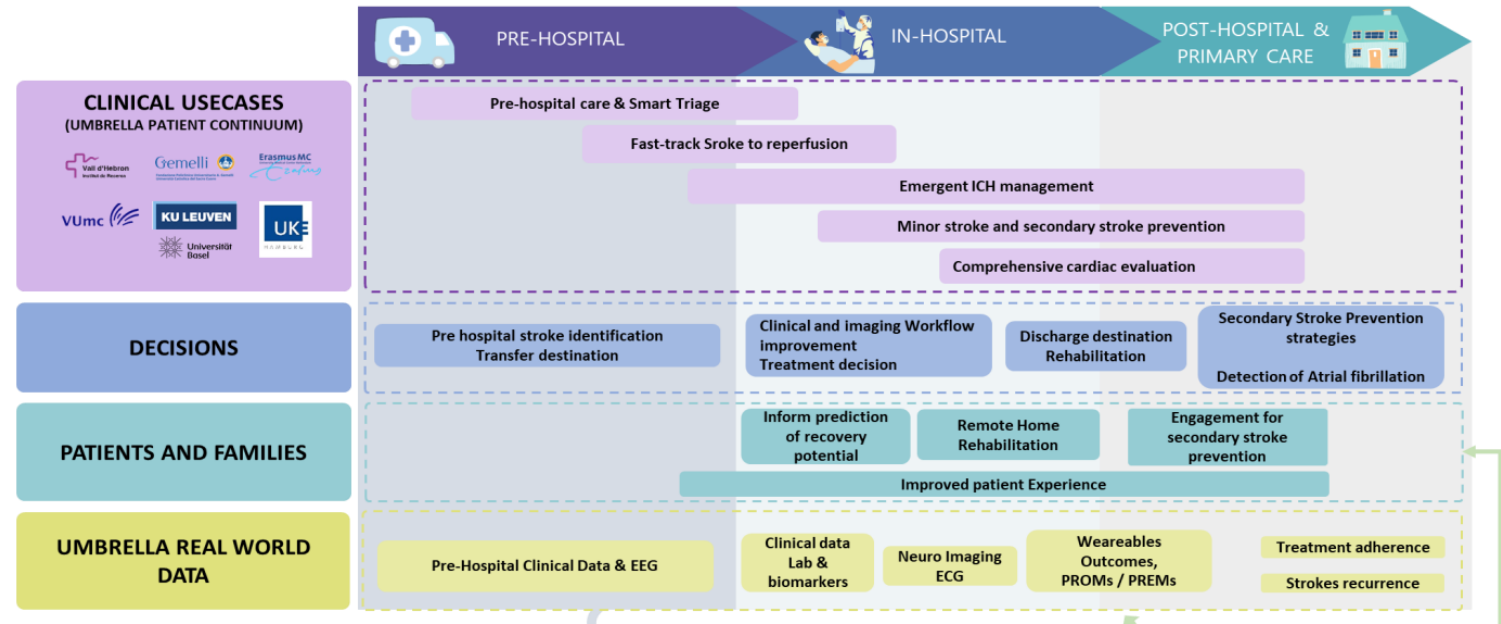
UMBRELLA

- **Federated Learning Platform**

- **Federated Learning Algorithms**

- **Model Server**

- **CAFEIN has the backend**

- **Siemens Cloud Platform has frontend**
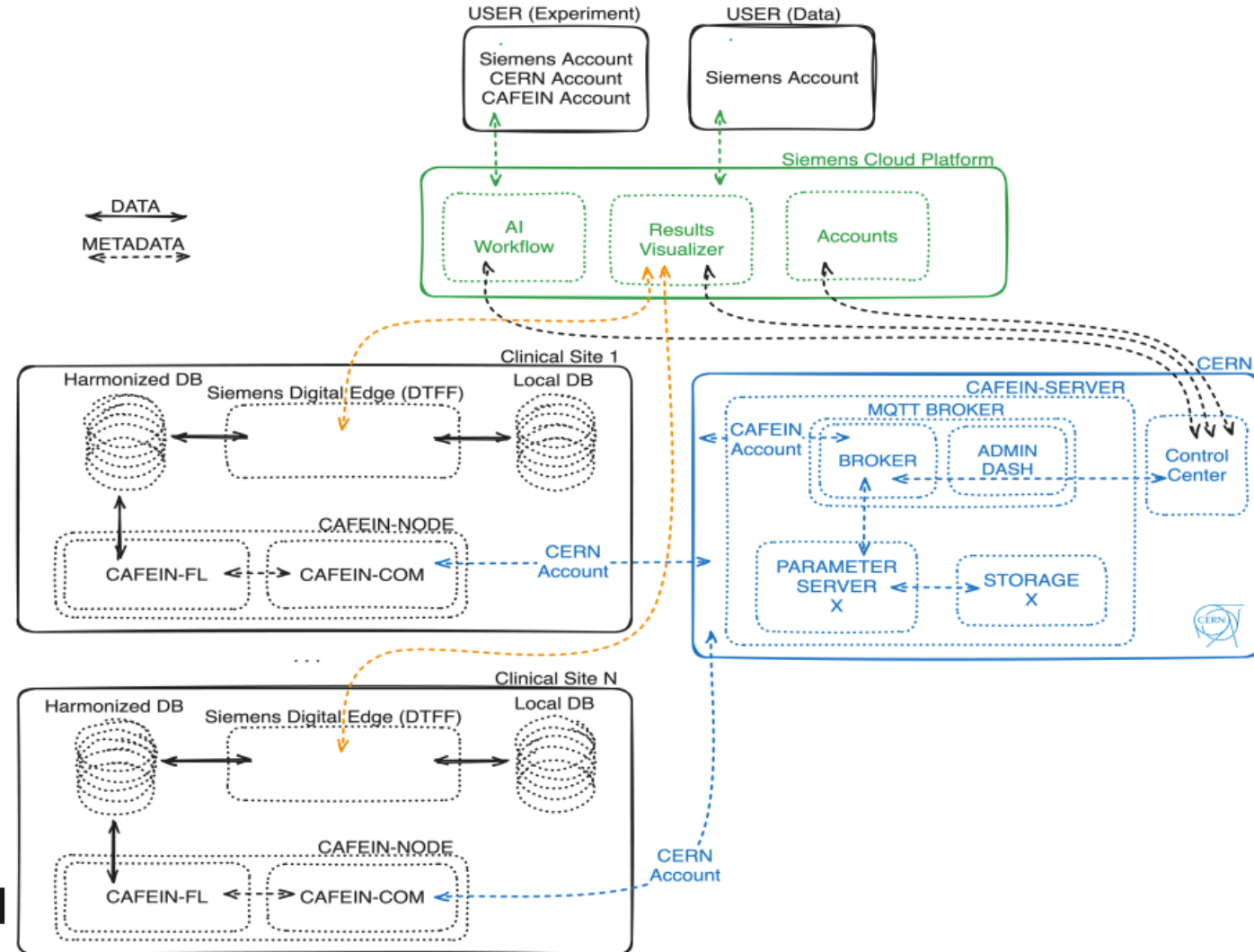
WHO Market Analysis Platform (MIP) for Prevention and Mitigation of Supply Chain Disruptions

- WHO's Operational Support & Logistics (OSL) is developing a Market Intelligence Platform ("Platform"), providing in-depth insight on market indicators, an alert system, market trends, demand clarity and supply chain capacities, and risk profiles associated with large-scale health emergencies.
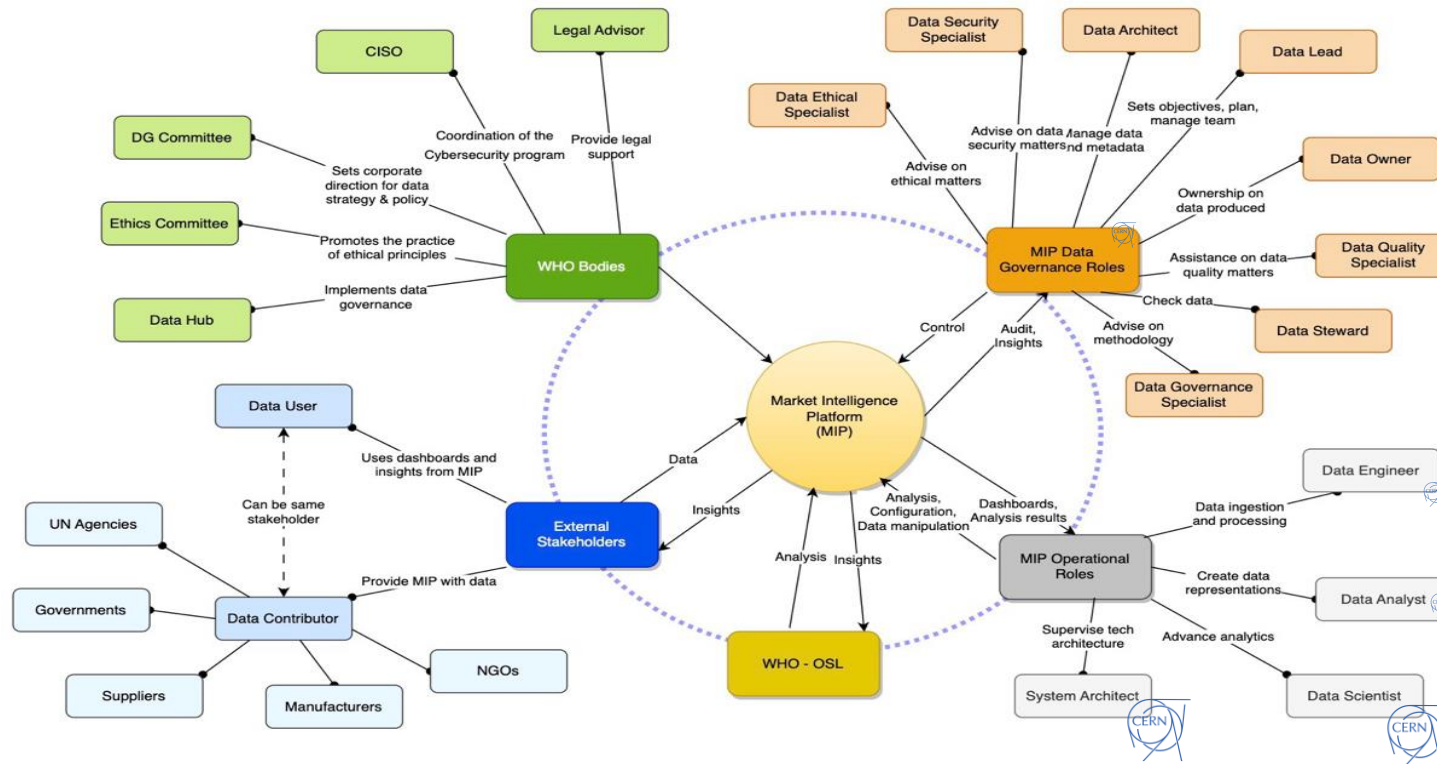
WHO Market Analysis Platform (MIP) for Prevention and Mitigation of Supply Chain Disruptions

CAFEIN – Federated Analytics

- Data analytics across multiple data owners without transferring any raw data.

CAFEIN – Federated Learning

- Training machine learning models across multiple data owners. Local models are trained using local data. Only local model updates are transferred. Local models are aggregated in a central server.

CAFEIN – Federated Inference

- Model prediction across multiple data owners
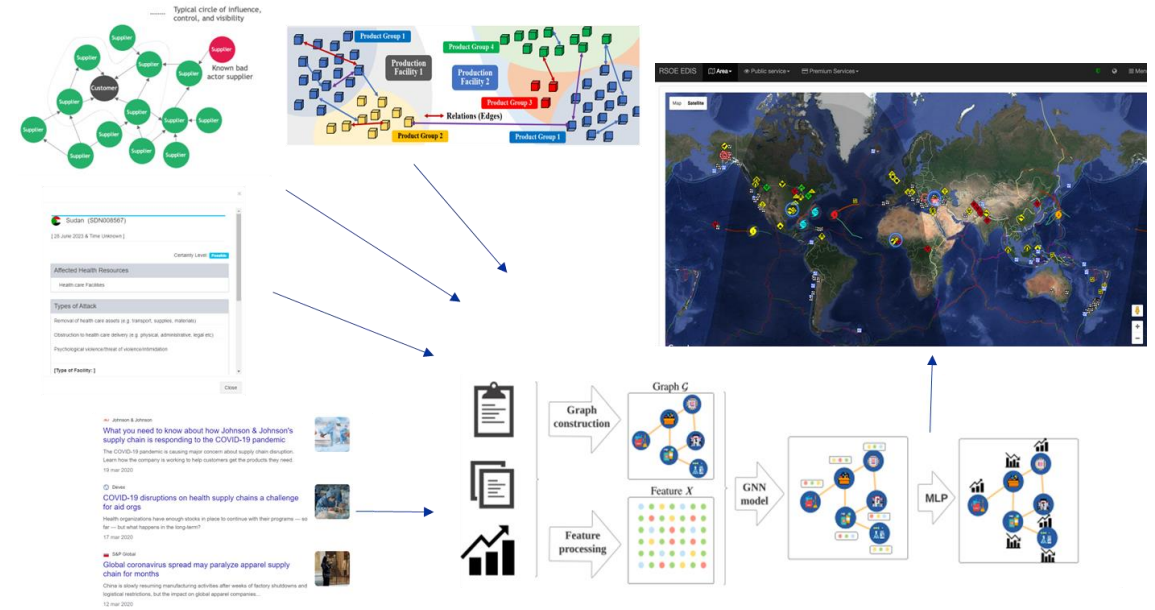
GNN for medical supply chain management

STELLA, Smart Technologies to Extend Lives with Linear Accelerators

- **Re-engineering the Next Generation of Medical Linear Accelerators for Use in Challenging Environments**

  - For nearly 60% of cancers, RT is most useful tool for cancer cure or palliation; inadequate supply of RT linear accelerators (LINACs).

  - 27.5 million new cancer diagnoses and 16.3 million projected cancer-related deaths worldwide in 2040. (WHO)

  - Current LINAC technology is complex, labor intensive, and high cost to acquire, install, operate and service.[1]

- **AI algorithms with Federated Learning to provide efficiently quality healthcare everywhere and at the same time enhance the robustness of the models with huge amounts of untapped data**

STELLA, Smart Technologies to Extend Lives with Linear Accelerators

Anomaly detection / predictive maintenance

- Collimators – detection of the temperature sensor failure – LSTM
- UPSs – detection of battery ageing – bi-directional LSTM
- Transformers – fault detection/RUL– autoencoders and random forest
- Electrical devices vs. Beam Dumps – RELIEF and Wrapper (Genetic + CS-SVM)
- Mining Dependencies of Systems and Components from Alarms Cascade – APRIORI
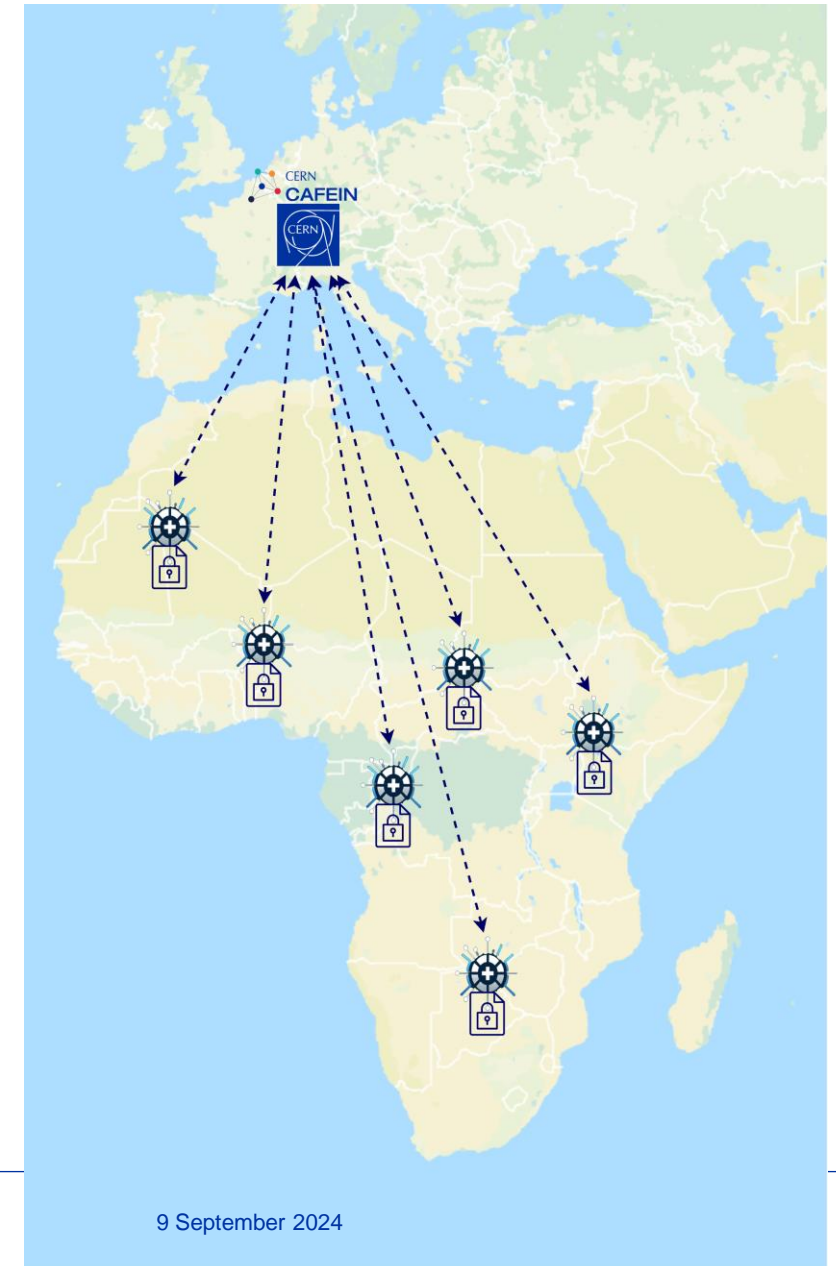- Automatic detection and classification of Welds – R-CNN

Operation

- For diagnosis and treatment

Diagnosis

- Tumor detection and segmentation from CT

Treatment

- Automated radiotherapy delivery

Pre-Conclusion

- **There are significant amounts of data/use cases for anomaly detection and predictive maintenance.**
  - How do you make it work/integrate it?
  - Build data pipelines/services for AI

- **There are several use cases for LLMs**
  - need to start and integrate with users/use-cases
  - backbone is operational for NLP for CRG
  - soon image/tables/SQL

- **Modeling/digital-twins of the accelerator complex can be built**
  - Diagnostics; advance control; design

- **Expertise in**
  - Federated learning
  - Algorithms: GNN, CNN, LLM
  - Application: Predictive maintenance, anomaly detection, medical image and signal analysis

Final Thoughts

- **Leveraging CERN technologies, expertise, and infrastructure, CAFEIN was developed.**

- **CAFEIN (and the team) are fully funded by EU projects and other public institutions.**

- **Synergies around CAFEIN, KT, and TE-CRG can be found.**
  - Resources and technical knowledge and expertise

KEY FACTS

CAFEIN - Federated network platform for the development and deployment of AI based analysis and prediction models

Submission Year
2019

Budget
135kCHF

Timeline
2019 - 2022

Funding Opportunities
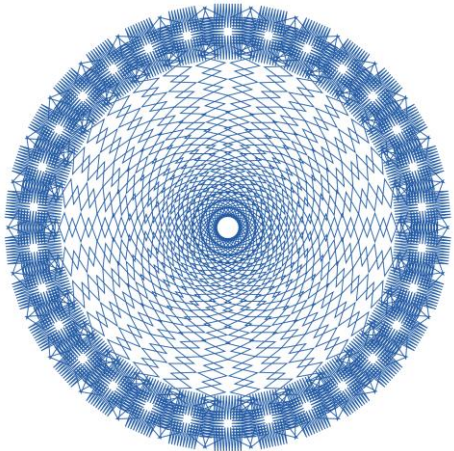CERN Medical Applications budget

Fields of Expertise
Simulations and Computing

CHF~3M

CERN'S IMPACT ON SOCIETY

One of the Management's top objectives for the next five-year period is to increase CERN impact on society, thereby boosting the Organization's visibility and consolidating the support of governments and the general public.

home.cern