Advancing patient care through health data engineering and trustworthy AI

**Vasileios C. Pezoulas** 1#, **Dimitrios I. Zaridis** 1, **Kosmas Apostolidis** 1, **Nikolaos S. Tachos** 1, **Kostas Marias** 2, **Manolis Tsiknakis** 2 and **Dimitrios I. Fotiadis** 1\*

1 Biomedical Research Institute (BRI) - Foundation for Research and Technology - Hellas (FORTH), Ioannina GR45110, Greece

2 Computational Biomedicine Laboratory (CBML), Foundation for Research and Technology - Hellas (FORTH), Heraklion GR70013, Greece

# Presenting author: Dr. Vasileios Pezoulas, email: bpezoulas@gmail.com.

\* Corresponding author: Prof. Dimitrios Fotiadis, email: fotiadis@uoi.gr.

**ABSTRACT**

In the rapidly evolving landscape of the healthcare domain, artificial intelligence (AI) and data engineering are game changers that can enable more precise, personalized, and efficient patient care, particularly in complex fields such as oncology. AI's ability to analyze vast amounts of data can create new paths for early diagnosis, tailored treatment plans, and predictive analytics that can significantly improve patient outcomes. Despite the benefits of AI in healthcare, significant challenges emerge to ensure that the developed AI systems are trustworthy. Examples of such challenges include the need for high-quality and interoperable data from multiple sources, the management of data privacy and AI risks, the compliance with regulations such as the General Data Protection Regulation (GDPR) in Europe, and the development of robust AI models that can generalize well across diverse populations [1]. To address these challenges, we present a multidimensional data engineering and trustworthy AI framework to advance AI in healthcare with respect to the principles of trustworthy AI. The proposed framework draws knowledge from: (i) the European Cancer Imaging Initiative (EUCAIM) [2] regarding the curation, semantic modeling and standardization of data from multiple sources (e.g. imaging, clinical) towards the establishment of federated health data spaces that can be interconnected and accessed through an advanced cloud based platform, (ii) the ProCAncer-I EU project [3] regarding the design, development and validation of advanced deep learning (DL) models specific to clinical scenarios (e.g. for prostate cancer diagnosis, clinical significant prostate cancer detection, prostate cancer characterization, treatment, and radiotherapy toxicity assessment), utilizing over 10,000 cases of multi-parametric MRI (magnetic resonance imaging) data to ensure that the AI models can be effectively generalized and deployed across diverse healthcare settings, and (iii) the FAITH EU project [4] regarding the assessment of the trustworthiness of the AI models throughout their lifecycle. This is achieved through an AI model passport which continuously monitors the AI model from its development to its deployment to identify multiple risks that are related to data and AI model bias (and fairness), data privacy concerns, lack of AI models’ explainability and interpretability, among others. Through this way the proposed multidimensional framework can support the development and deployment of AI models that are not only more accurate but also more reliable in real-world clinical settings to ensure that the AI models are trustworthy and can be reliably used in the clinical decision-making process.

**Keywords:** health data engineering, trustworthy AI, federated data spaces, data curation, data standardization.

**REFERENCES**

1. Huerta, E. A., et al. "FAIR for AI: An interdisciplinary and international community building perspective." Scientific data 10.1 (2023): 487.
2. <https://www.egi.eu/project/eucaim/>
3. <https://www.procancer-i.eu/>
4. <https://faith-ec-project.eu/>

**14th Scientific FORTH Retreat, October 2024**