How Smart Tech Makes Health Records Smarter (PREPARE Solutions for Better Healthcare)

Erik Novak[1], *Erik Calcina*[1], *Tanja Zdolšek Draksler*[1]

[1]Jožef Stefan Institute, Ljubljana, Slovenia

Background: The PREPARE project, "Personalized Rehabilitation via Novel AI Patient Stratification Strategies" (Grant Agreement No: 101080288, 2023-2027), is funded under the Horizon Europe initiative. It aims to tackle the challenges of rehabilitation care for patients with chronic, especially non-communicable, diseases. The highly personalized nature of each patient's condition makes prediction tools and patient stratification unreliable. The project addresses these challenges through a combination of innovative clinical, social-behavioral, and public health research, along with advanced AI and data science methods. The project involves 20 partners from nine countries across Europe, including universities, research centers, and healthcare providers.

PREPARE Technologies: extracting information from unstructured healthcare data: Within the PREPARE project, we are developing machine learning (ML) techniques to process nine extensive patient datasets. These techniques include methods for extracting relevant medical information from unstructured healthcare data, such as comments and reports. We are designing Named Entity Extraction (NER) models to identify and retrieve relevant named entities from text, such as people, locations, and organizations. The NER models in PREPARE focus on extracting medical entities, such as drugs, events, measurements, and treatments. We are also experimenting with Large Language Models (LLMs) that can generate responses containing the relevant extracted information. Due to the multilingual nature of the patient datasets, we are developing models that support various languages. Additionally, we are creating methods to monitor the performance of these extraction techniques. Since the extracted values are used in patient stratification techniques, it is crucial that the healthcare data is accurately extracted.

Conclusions: The PREPARE project is ongoing. The extraction techniques are still in development and will be applied to various patient stratification-related healthcare datasets. The extracted information will enhance the existing structured datasets available to the healthcare provider partners. Furthermore, information extraction is part of a larger pipeline to prepare patient data, which will be converted into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) standard. This will enable us to develop prediction and stratification machine learning strategies for rehabilitation data, which will be validated through nine demonstration pilots.