EURASIA CONFERENCES
3rd World Conference on
ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND DATA SCIENCE
May 23, 2024 | Rome, Italy
https://doi.org/10.62422/978-81-970290-9-7-004

Automatic extraction of SMPC document for IDMP data model construction using Open-Source base Model LLM RAG: A benchmark for Pharmaceutical Regulatory Affairs

Florian PEREME
Digital Innovation Leader - Research & Innovation Dept, Product Life Group, Paris France

The pharmaceutical industry is on the cusp of a deep transformative change that would lead to a more digitalized industry. One of the markers of this upcoming changes is the implementation of IDMP (Identification of medical drug product) who is the digitalization in a structured and harmonized data model of all information related to a drug product. Historically those data are present in hard files such as PDFs, and exchanged between Industries and Health Authorities, requiring Human Work Forces on both sides to produce and interpret the data valuable content from this file. The IDMP ambitions is to improve this process by making possible direct data exchange between Industries and Health Authorities. This deep Changes comes with a huge challenge to handle legacy content for Industries in order to rebuild the IDMP structure from existing SMPC document. Our contribution proposes to take advantage of the new LLM Solutions and their RAG (Retrieval Augmented Generation) to perform a benchmark of existing LLM solution and evaluate their abilities to proceed to relevant data extraction an rebuild IDMP base data content.

Biography:

Ph.D. in Engineering, with exêrience in Services Company, across multiple field for Research and Innovation with Digital Application. More specifically focused on native cloud solution and Machine Learning applications.