

AccGPT: A Vision for AI Assistance at CERN's Accelerator Control and Beyond

AccGPT is an on-going project to integrate AI into various levels of operations CERN's operations, particularly in the domain of particle accelerator control. The goal is to embed AI assistants in critical areas: aiding control room operations for managing accelerators, assisting in coding for development purposes, and enhancing the effectiveness of documentation and knowledge retrieval. These AI systems will also streamline answering frequently asked questions, thereby increasing operational and onboarding efficiency.

AccGPT has already made significant strides with a robust prototype capable of retrieving knowledge from CERN's internal documentation. Efforts are also underway to integrate AI into coding assistance, targeting the organization's internal code repositories. This endeavor is pivotal in leveraging AI to ultimately increase productivity.

In parallel, a collaborative framework has been established at CERN, encompassing the AccGPT project and other AI chatbot use cases, including those involving access-restricted data. The primary objective of this collaboration is to share hardware resources and common building blocks, fostering synergies and promoting collective advancements in the field.

With the dedication of more GPU hardware resources in the future, we aim to further improve the accuracy with training or fine-tuning open-source models to our dedicated datasets.

This contribution will describe the technical choices for AccGPT, the challenges encountered, and the results obtained. Insights into the on-going work and future outlook will also be covered.

Primary author: REHM, Florian (CERN)

Co-authors: GUIJARRO, Juan Manuel; SOUFFLET, Nathan; KAIN, Verena

Presenter: REHM, Florian (CERN)

Track Classification: Automated text and code generation