Contribution ID: **137**   Type: **Presentation**

Utilizing large language models with free and open source software in EFSS while protecting digital sovereignty for Universities.

Wednesday 13 March 2024 14:15 (15 minutes)

The recent media buzz around so called "Artificial Intelligence", divides users of IT systems into two polarized groups. One group is vocal about wanting to use these new tools and points out that these features are missing from offerings outside of IT giants like OpenAI/Microsoft and Google. The other group is worried about copy right and privacy issues that arise from having all your files and communications scanned by large corporations for generating and commercializing large language models (LLM) and other machine learning (ML) tools.

Luckily, as competent engineers, we can appease both of these groups at the same time using free and open source tools and ethically sourced models running on your own hardware. In this talk I will present how Local AI[0] can be integrated with Nextclouds ML tools[1,2,3]. Having access to GPU resources is helpful, but not necessary for decent results.

You would be forgiven for thinking that large enterprises such as Google and OpenAI have access to resources not available to the general public, and while that is certainly true, it is true for marketing resources more than anything else. In fact machine learning has long been the domain of academia and the free and open source movement, and the contribution of the large corporations in the space is mostly doing the work at scale and defining expectations of users.

There are already mature projects available for running LLM:s on commodity hardware with the same API endpoints as OpenAI presents. This opens up the possibility to integrate with the recent tools that Nextcloud have developed for summarizing text, writing headlines and more (also integrated in text and mail apps for example).

1. https://localai.io
2. https://github.com/nextcloud/integration_openai
3. https://github.com/nextcloud/assistant
4. https://github.com/nextcloud/mail

Primary author: NORDIN, Micke (SUNET)

Co-authors: Mr ANDERSSON, Magnus (SUNET); Mr FREITAG, Richard (SUNET)

Presenter: NORDIN, Micke (SUNET)

Session Classification: Technology Bricks: advanced integration

Track Classification: Main sessions: Technology & Research