



## Updating confsearch with language models

The project aims to integrate Large Language Models (LLMs) into Confsearch to automate the verification and improvement of conference data, thus enhancing data accuracy and reducing manual intervention. Using LLMs in the web-scraping process, the system will automatically update and correct conference details, with provisions for manual user overrides. The project will include a documentation process to ensure future maintainability, covering the architecture, code, and other guidelines.

The exact scope of the project is not yet fixed and is open for discussion; we are looking for motivated students who are excited to work on enhancing a tool used by many researchers.

**Requirements:** Programming skills (Python, C/C++, ...) and knowledge of web development. Prior experience working with large language models and web scraping is an advantage.

We will have weekly meetings to address questions together, discuss progress, and think about future ideas.

### Contact

In a few short sentences, please tell us why you are interested in the project and about your coding and machine learning background (i.e., your own projects or courses).

- Andreas Plesner: [aplesner@ethz.ch](mailto:aplesner@ethz.ch), ETZ G95