

## Research project for the position financed by European Social Fund Plus

### *Title:* Language technologies for low-resource languages

*Description:* In the domain of artificial intelligence, large language models have recently led to a boost in the development of language technologies which recognize and produce language both in oral and written form (e.g., chatGPT). However, a current challenge is the development of systems able to deal with 'small' languages (i.e. languages with few speakers) or non-standard languages (e.g. regiolects or dialects), for which not many resources are available which can be used to train the AI. In order to extend large language models to include low-resource languages, experts from the field of linguistics are needed which are able to define the needs of speech communities with respect to language technologies, can prepare datasets for training, and can contribute themselves to the creation of tools.

The goal of this project is to form such experts at the PhD level by providing a solid basis in the domain of linguistic analysis and sociolinguistics, as well as in domains connected to computational linguistics.

The region of South-Tyrol provides a perfect testing ground for language technologies of this type, with the presence of 'small' languages such as Ladin and its high density of non-standard Tyrolean dialects.

*Supervisor:* Prof. Alessandro Vietti, L-LIN/01