

## Lab 2 Pre-Lab Tasks: word2vec and fasttext embeddings

---

To avoid the delays caused by slow downloading that we had in the first lab, please complete the following setup tasks **before** class on Thursday:

1. Create a new python project and a virtual environment for it in the usual way.
2. Install the following packages:

- gensim
- compress-fasttext
- matplotlib
- numpy
- scikit-learn

**There is currently a small problem with the gensim package.** It relies on several functions in the *scipy* package that have been recently deprecated. Until the gensim package is fixed, you need to uninstall scipy 1.13 (which was installed along with gensim, and where the functions have been removed), and install scipy 1.11.4 (an older version where the functions are still available).

3. Download the following embeddings and save them in your project directory:
  - [Google News Vectors -English](#). The download link is about 2/3 down the page, in section "Pre-trained word and phrase vectors". The file is named **GoogleNews-vectors-negative300.bin.gz**. Do not unzip the file, it can be loaded as-is.
  - [FastText Compressed Binary Vectors - German](#). Download **fasttext-de-mini**.