

## Project 4

# SäxyLLM: Saxon LLM construction

Sächsisch is sexy, but current open models don't speak it. Build an LLM that speaks Saxonian by default.

## Additional Materials

Here are links that you can use for your training dataset. However, they are not mandatory, so feel free to use anything you find in saxonian.

1. Text resources:
  - A large set of poems:  
<https://www.sachsen-lese.de/streifzuege/mundartliches/>
  - 19 encyclopedic articles:  
[https://incubator.wikimedia.org/wiki/Wp/sxu/Main\\_Page](https://incubator.wikimedia.org/wiki/Wp/sxu/Main_Page)
  - 9 poems: <https://de.wikisource.org/wiki/Kategorie:Erzgebirgisch>
  - Various short phrases:  
<https://www.scribd.com/document/408398334/alltagsdeutsch-sachsisch>
2. Rules (for rule based synthesizing):
  - [https://praxistipps.focus.de/saechsisch-lernen-diese-basics-sollten-sie-wissen\\_157089](https://praxistipps.focus.de/saechsisch-lernen-diese-basics-sollten-sie-wissen_157089)
3. Audio files need transcription:
  - <https://www.regionalsprache.de/Audio/Catalogue.aspx>
    - (search by location, e.g., "Dresden")
  - <https://dsa.info/sachsen/>
4. Other audio files:
  - [https://commons.wikimedia.org/wiki/Category:Obers%C3%A4chsisch?use\\_lang=de](https://commons.wikimedia.org/wiki/Category:Obers%C3%A4chsisch?use_lang=de)
5. Long audio book for 1 €, perhaps this would be the biggest resource:
  - <https://www.amazon.de/Winnetou-auf-s%C3%A4chsisch-Karl-May/dp/>