

Student Assistant (Studentische Hilfskraft) for Project ECALP

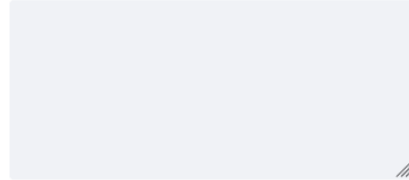
Trustworthy Human Language Technologies (<https://www.trusthlt.org/>)

[TrustHLT](#) is looking for a **Student Assistant** to help creating a novel multi-modal corpus to advance research in the field of **Artificial Intelligence in Legal Proceedings**. The corpus should consist of videos of court hearings of the European Court of Human Rights along with annotations and transcriptions of the oral hearings.

What is this about?



Attendants:



- around 280 videos
- without subtitles
- with multiple speakers and languages
- automatically generated transcripts with Whisper (OpenAI)

Transcript

```
<<631.68;634.16;SPEAKER_10;Government;fr>>  
On peut, bien sûr, les porter.  
  
<<634.16;640.84;SPEAKER_10;Government;fr>>  
Les autres prescriptions, par exemple, ne pas s'attarder habituellement dans des bistros et des  
tripots.  
  
<<641.12;644.84;SPEAKER_10;Government;fr>>  
Je crois qu'il ne soit pas des choses très graves.  
  
<<645.0;656.04;SPEAKER_10;Government;fr>>  
C'est mieux de rentrer les soirs avant 6 heures du soir et ne pas sortir avant 6 heures du matin.  
  
<<657.36;659.72;SPEAKER_10;Government;fr>>  
Ce ne sont pas de grandes restrictions.  
  
<<660.12;669.04;SPEAKER_10;Government;fr>>  
Donc, il est évident que les prescriptions considérées dans leur ensemble ne entraînent pas une  
prévision de la liberté personnelle.
```

THE TASK:

- ❖ Compiling a corpus of court hearing transcripts from the ECHR (<https://echr.coe.int/Pages/home.aspx?p=hearings>)
- ❖ Annotating timestamps, speakers and identifying spoken languages
- ❖ Reading the automatically created transcripts and manually correcting mistakes in the transcripts to obtain a gold standard corpus
- ❖ Future tasks involve applying SOTA models for argument mining and developing new methods to tackle argumentation in legal proceedings with this new corpus!

REQUIRED SKILLS:

- ❖ Some prior experience in python (pytorch, huggingface, pandas...) is helpful
- ❖ **Very Good** language skills (**French required**, English helpful)
 - You need to be able to verify and correct the transcripts manually
 - Additional language skills can also be helpful! The official languages of the court are English and French.
- ❖ Motivation to work on a challenging task in a largely unexplored field

FORMALITIES:

- ❖ Studentische Hilfskraft at TrustHLT (TU Darmstadt)
- ❖ Up to 40 hours monthly (flexible)
- ❖ Start as soon as possible

Interested? Just send a
short application to:

lena.held@tu-darmstadt.de