ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

Prof. R. Wattenhofer

# Use LLMs to rank movies

Have you ever watched a bad movie? How did you know it was bad? Maybe the visuals or the acting was just b-tier, or maybe it was the writing? Did the character dialogue just seem stiff and predictable? If you ever had this experience, you're not alone!

In this thesis, we will get to the bottom of what it means to have a bad script, measured as a function of predictability. We will rank movies and TV shows based on their screenplay predictability and see how well this correlates with their IMDB/Rotten Tomatoes score. To this end we will aggregate and clean a large corpus of publicly available screenplays and evaluate a set of LLMs on them. Depending on our findings we will try to improve bad screenplays through an unusual use of LLMs.

**Requirements:** Strong Python programming skills. Preferably with interests in data scraping and cleaning, as well as large language modelling. Ideally interested in writing up a conference paper.

Weekly meetings will be scheduled to address questions, discuss progress, and brainstorm future ideas.

## Contact

- Luca Lanzendörfer : lanzendoerfer@ethz.ch, ETZ G93
- Frédéric Berdoz : fberdoz@ethz.ch, ETZ G60.1