**Karlsruher Institut für Technologie
Institut für Angewandte Informatik und
Formale Beschreibungsverfahren (AIFB)
Forschungsgruppe Web Science**
M.Sc. Shuzhou Yuan
shuzhou.yuan@kit.edu

Call for Bachelor/Master Thesis:
„Using Visual Text Representation for Robust Text Classification "

Background

With the spread use of social media, people use misspelling, typos or add specific word [1] to post malicious content deliberately to evade the censor of the platform. Human can still understand the meaning of the text, because we not only read the text but also see it. Inspired by the language processing of human being, we propose to utilize visual text representation [2] to strengthen the robustness of text classification tasks, e.g.: hate speech classification.

Source: If you want to know more about me, just give me a call.
With noise: 1f u w0nt t0 kn0vv mor3 ab0ut me, ju5t give me a 👍📞.

Figure 1: Example of text with noise. Human can still understand it, but traditional text classifiers usually fail to recognize it.

Goal

The goal is to investigate visual text representation and strengthen the robustness of models to against attack or noise for text classification tasks. This includes the following steps:
1. Investigating different types of attack or noise and adding them to existing datasets.
2. Creating visual text datasets by rendering text as images.
3. Comparing the performance of traditional text classifiers e.g.: BERT, with visual re-inforced models.

The details and contents of the Bachelor's/Master's thesis are flexible. The student will get the chance to publish the work as a research paper.

What are the prerequisites?

- Solid programming skills (e.g. Python).
- Strong interest in natural language processing or computer vision.
- Experience in pre-trained language models or HuggingFace library is a plus.

[1] https://arxiv.org/pdf/1808.09115.pdf
[2] https://aclanthology.org/2021.emnlp-main.576.pdf

Please send your requests with a transcript of records and a short CV to:
**Shuzhou Yuan
shuzhou.yuan@kit.edu**