Coling-UniA at GermEval 2025 Shared Task on Candy Speech Detection: Retrieval Augmented Generation for Identifying Expressions of Positive Attitudes in German YouTube Comments

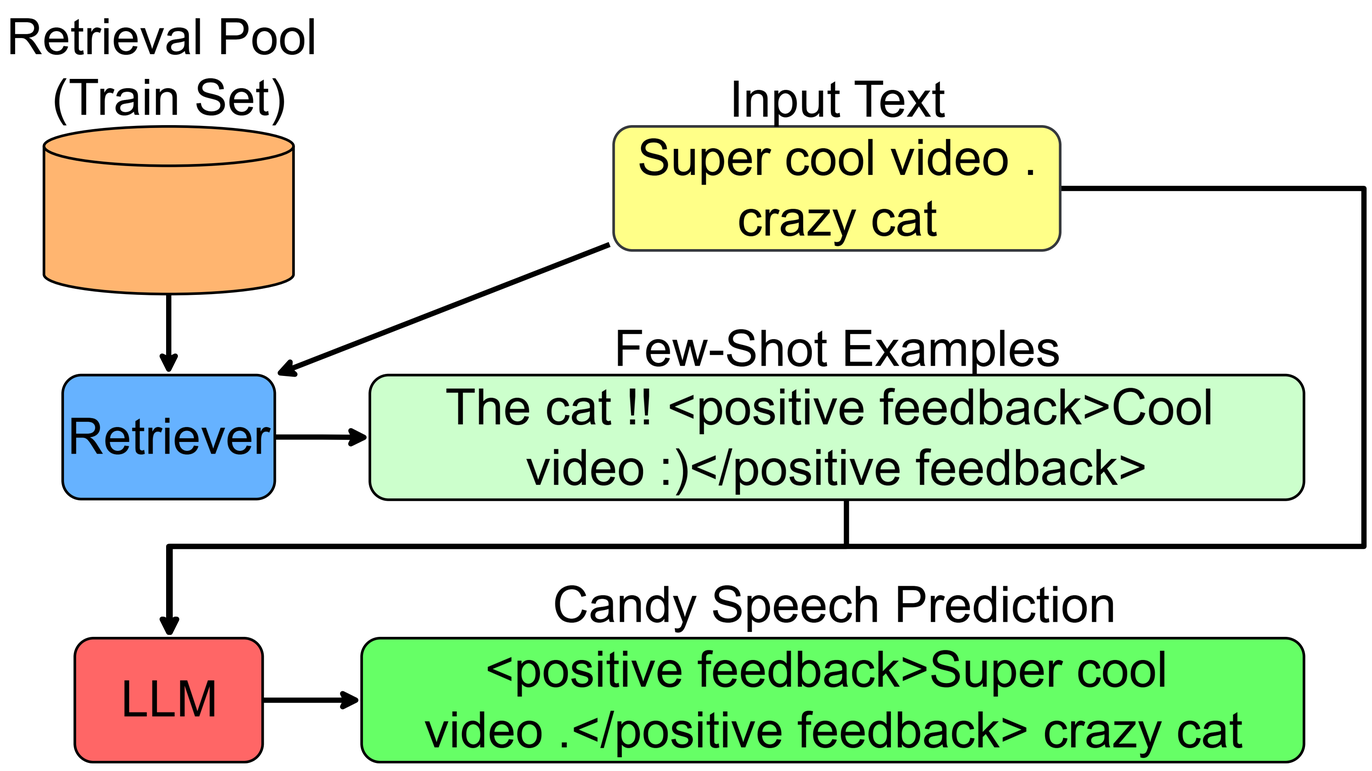
Georg Hofmann Annemarie Friedrich
Faculty of Applied Computer Science, University of Augsburg, Germany

Task Description

- Identify the exact spans of candy speech expressions in German YouTube comments and classify each into one of ten pre-defined categories.

Proposed Approach

- Use Large Language Models (LLMs) and a dense retriever in a Retrieval-Augmented Generation pipeline, exploring two target output formats: an XML-style and a CoNLL-style format

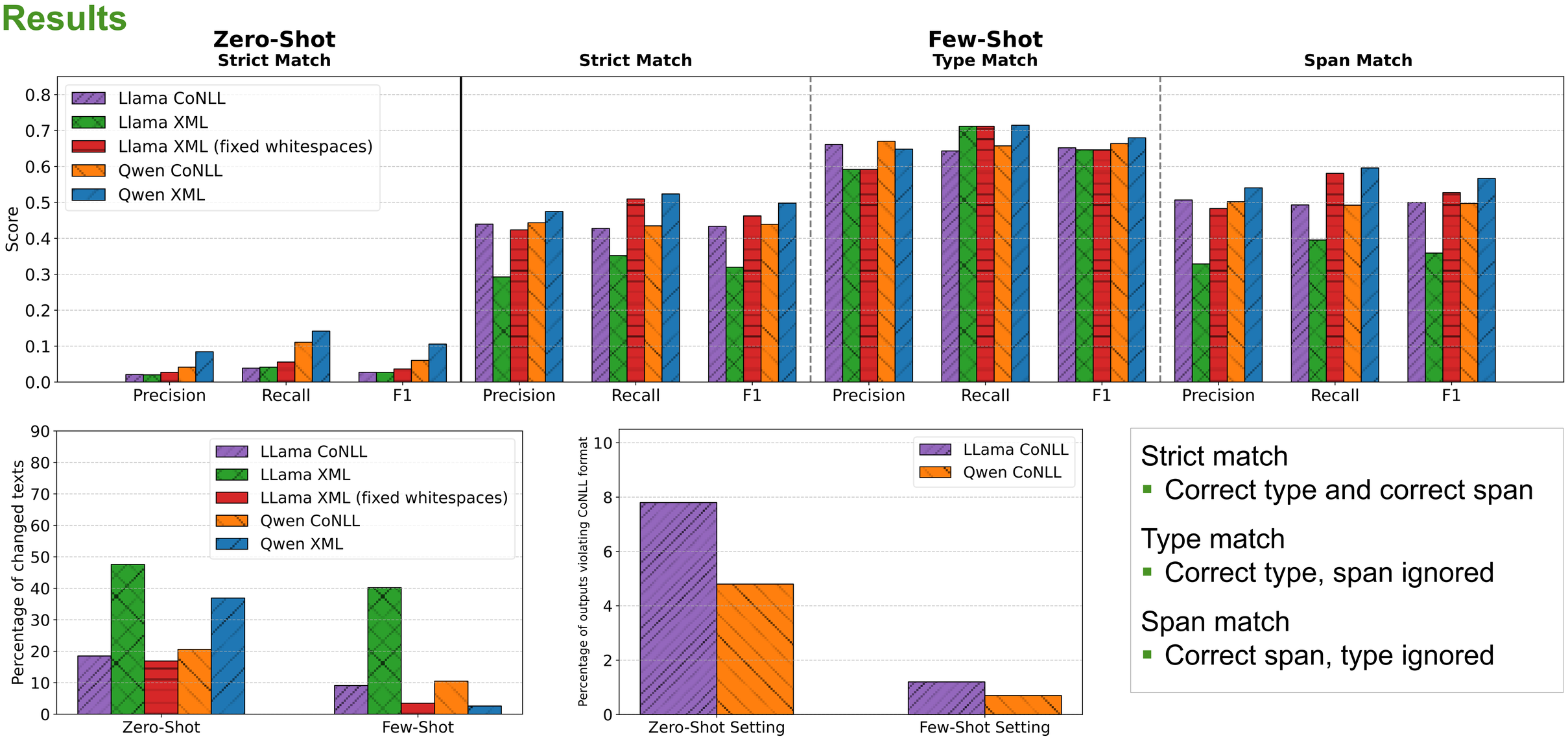


- LLMs used: Llama-3.3-70B-Instruct and Qwen2.5-72B-Instruct
- Llama with XML-style format often removed whitespaces, causing span errors; fixed via a simple postprocessing step to restore spacing
- Evaluate in zero-shot and few-shot settings

CoNLL-Style Format

Super	B-positive feedback
cool	I-positive feedback
video	I-positive feedback
.	I-positive feedback
crazy	O
cat	O

Results



Strict match
- Correct type and correct span

Type match
- Correct type, span ignored

Span match
- Correct span, type ignored

- LLMs occasionally alter the original input text in both prompt formats, for example, by correcting spelling or making small grammatical adjustments and other times by making unrelated changes, which can also negatively affect span predictions.
- LLMs sometimes fail to follow the CoNLL format, producing malformed outputs that could not be evaluated.

Link to repository

Georg Hofmann, Annemarie Friedrich
Universität Augsburg, Contact: georg.hofmann@uni-a.de

KONVENS 2025
9 – 12 September, 2025
Hildesheim, Germany