

# FauxPas-EAI Data and Evaluation Issues Analysis

For FauxPas-EAI we encountered both problems in the data and in the automatic evaluation. Here we show an example for both of them.

## Data Issues

We removed some stories from the dataset because they contain inaccuracies in the context or in the questions. The following example displays an error in the question, where Simon and Robert are placed in swapped order:

"Context": **Simon bought Robert a toy car** for his birthday. A few months later, they were playing with it, and Simon dropped it. 'Don't worry,' said Robert, 'it was only an accident'. "  
"Question": **Did Simon know Robert had given him the toy car** for his birthday? Answer with 'Yes' or 'No' only, without explanations. In case of doubt, answer according to the most probable answer."

## Evaluation Issues

Here we show an example of the inadequacy of the Levenshtein distance for the FauxPas-EAI evaluation. The model answers correctly, but the metric evaluates the response as wrong due to the strict matching requirements.:

"Context": "Robert had just started at a new school. He said to his new friend, Andrew, 'My Mum is **a dinner lady at this school**.' Then Claire came over and said, 'I hate dinner ladies. They're horrible.' 'Do you want to come and play rounders?' Andrew asked Claire. 'No' she replied 'I'm not feeling very well.'"  
"Question": "What job does Robert's Mum do?"  
"Ground truth answer": "Robert's Mum is **a dinner lady**."  
"Model answer": "She is **a dinner lady at the school**."