

# **REPORT**

## **THOUGHT PROCESS**

My goal going into this task was to find a good metric to evaluate code suggestions. I was also excited to use the Hugging Face transformers, as I always enjoy using them.

## **FINDINGS**

I found that the simple metric of just checking if the suggestion is a substring of the correct result, or if the correct result is a substring of the suggestion worked best, even better than chrF. Of course, there are even better metrics, but this one did a good job.

## **LEARNINGS**

Through this task I learnt that evaluating the results of code completion is harder than it looks as only subtle differences can make a suggestion completely wrong.