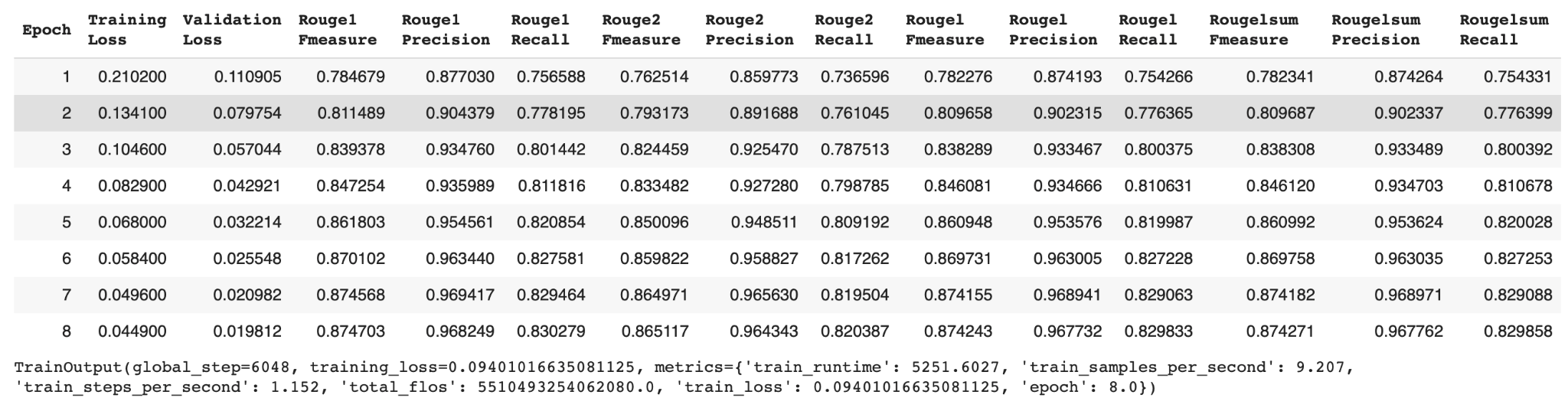
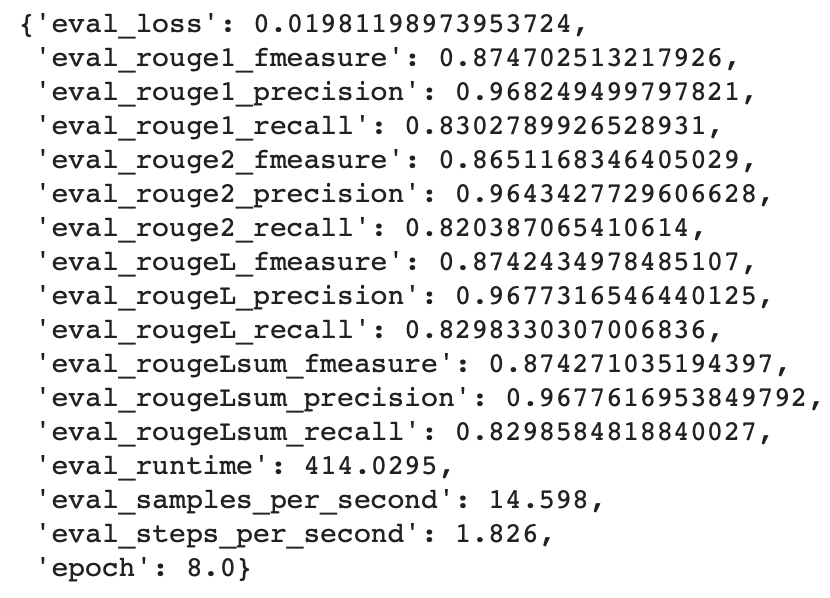
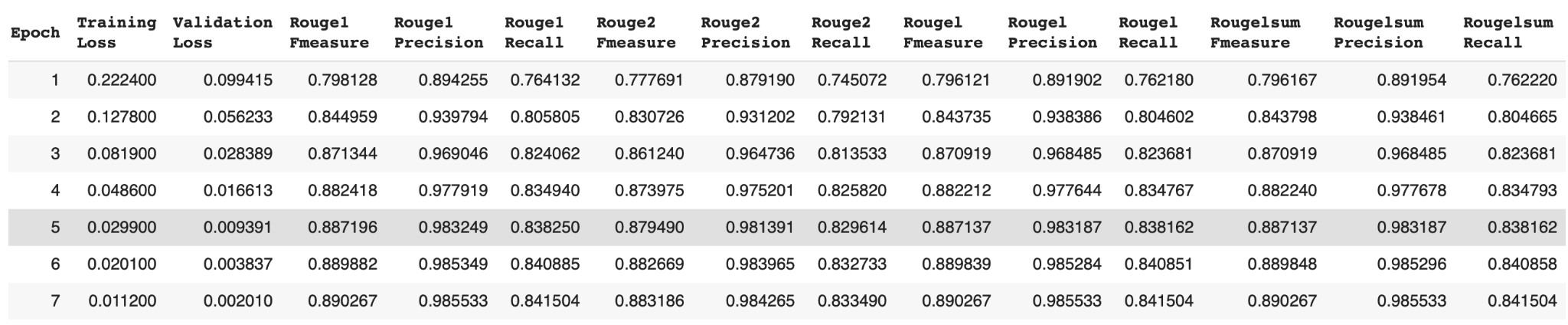
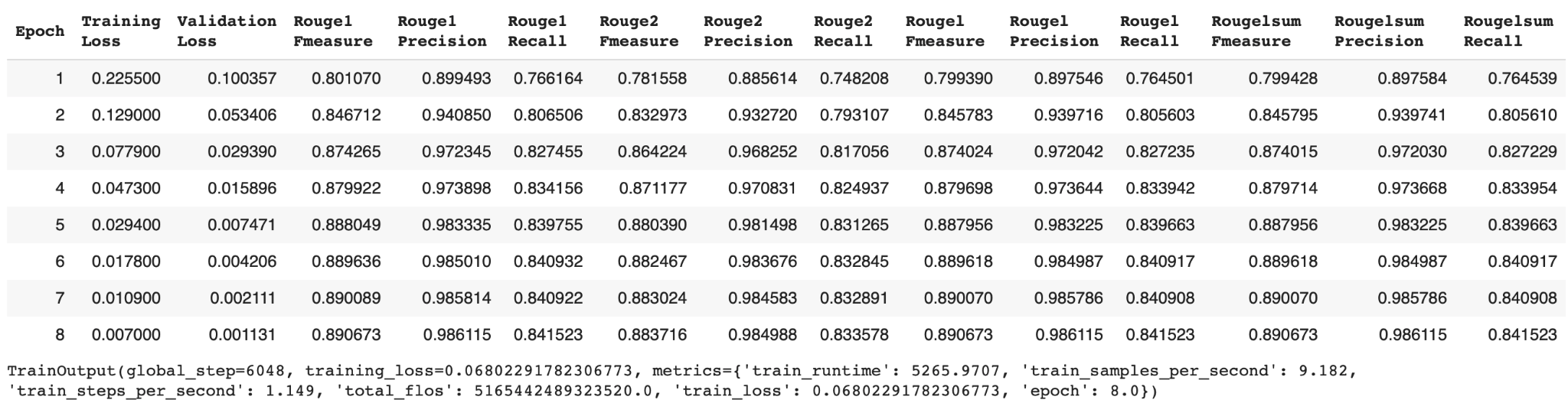
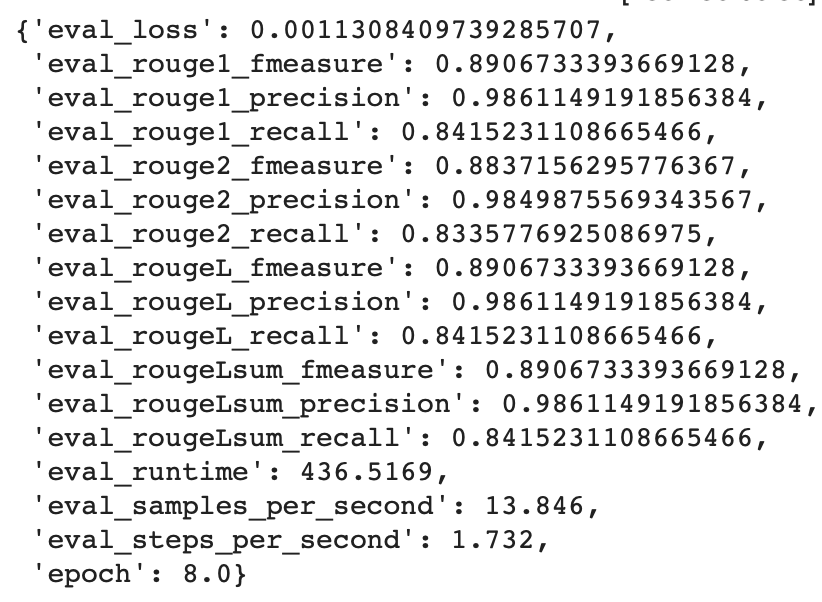
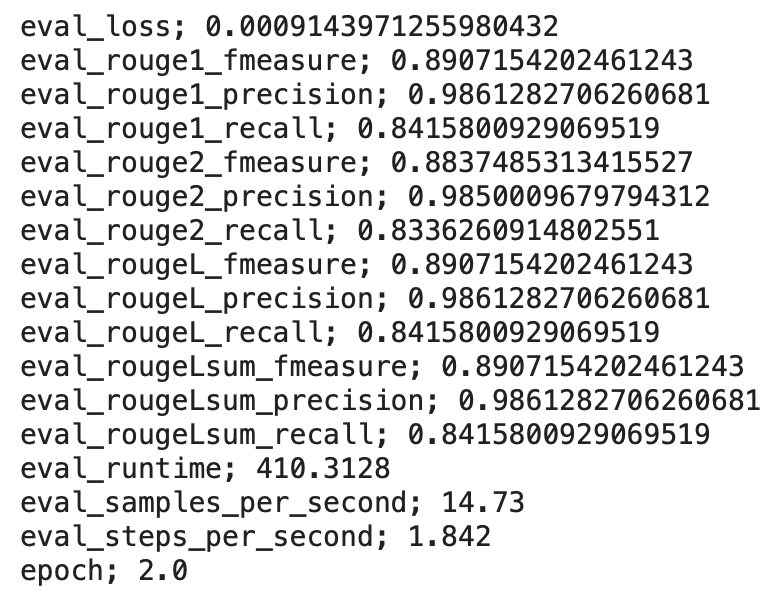
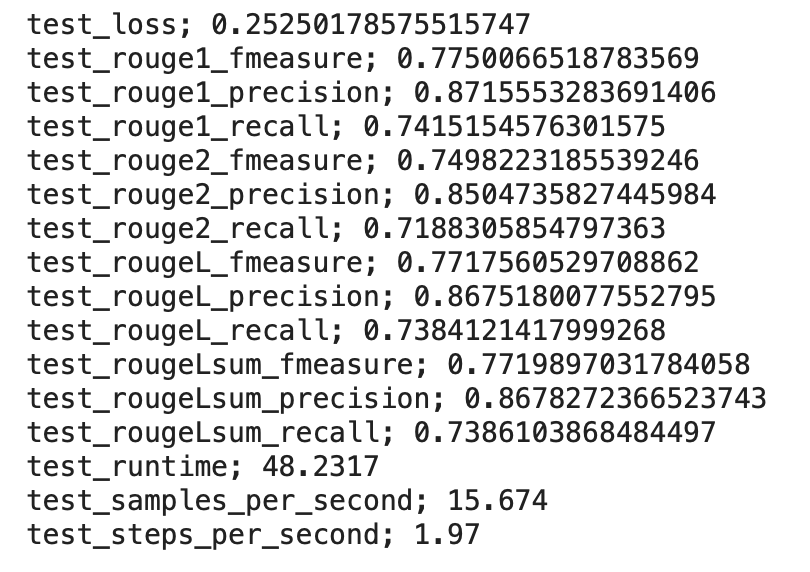
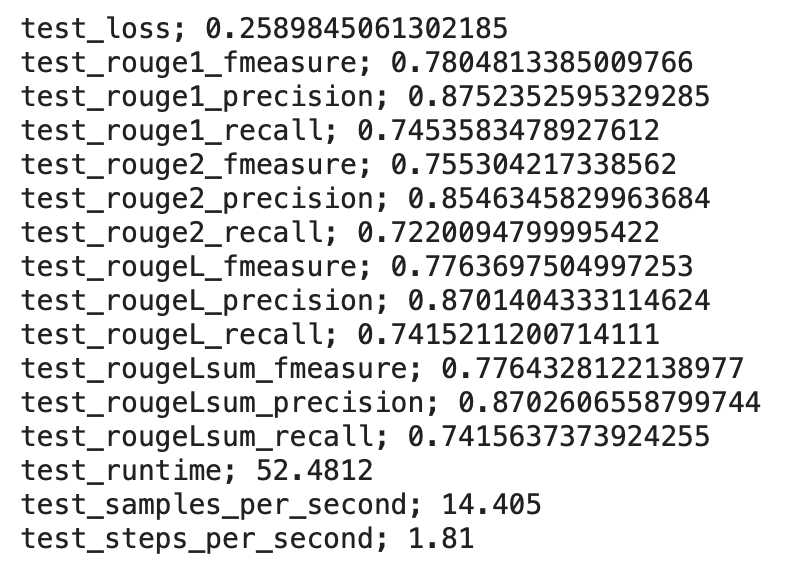
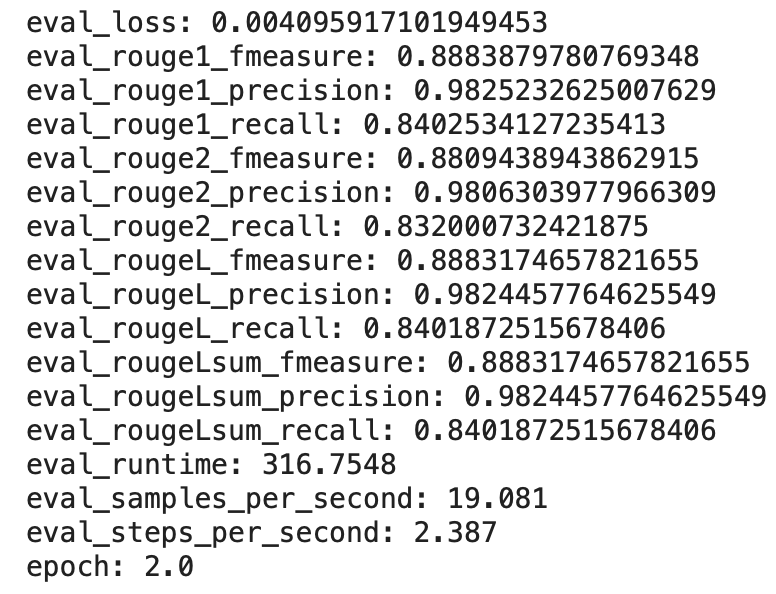
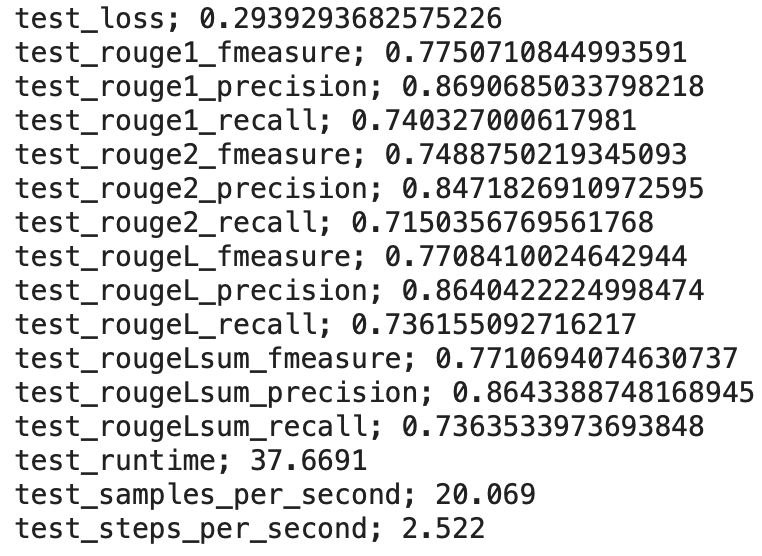
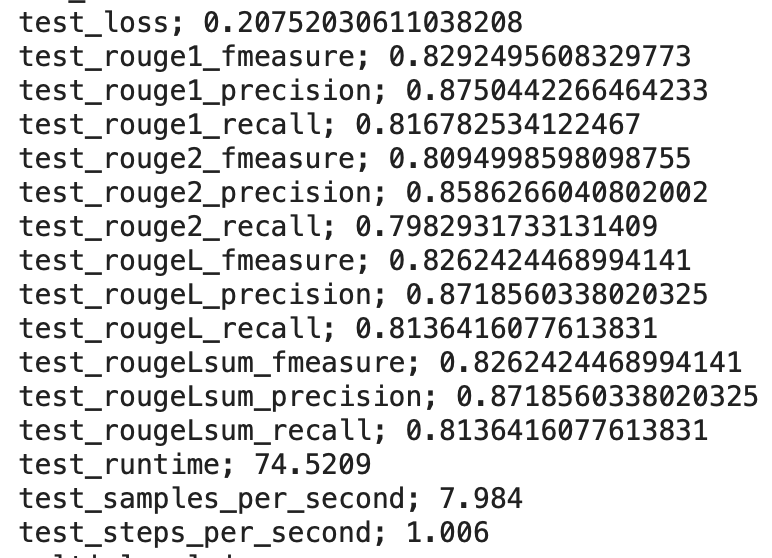
experimental results for T5

* t5-finetune-1: QA setup with LR 5.6e-5
  + epoch results:  
    
  + evaluation results:  
    
* t5-finetune-2: QA setup with LR 5e-4
  + epoch results:  
    
* t5-finetune-3: no QA setup with LR 5e-4
  + epoch results:  
    
  + evaluation results:  
    
* T5 trained as much as possible for good results → no QA formatting
  + evaluation results:  
    
* corrected T5 highly trained with no qa formatting:
  + evaluation results:  
    
* correct T5 trained with QA formatting:
  + evaluation:  
    
* BART with high learning rate and no QA setup:
  + evaluation results:  
    
* fixed BART metrics
  + evaluation results:  
    

# one claim vs multiple claims

* eval result for tweets with one claim:  
  
* eval result for tweets with multiple claims:  
  