SMS Spam Detection using Large Language Models

George Kyriopoulos

Objective

The objective of this project was to develop an SMS spam detector using Large Language Models (LLMs) and subsequently conducting a comparison of these models.

The models used were RoBerta, Bert, DistilBert and AlBert.

Hugging Face served as the primary framework.

Dataset

The SMS Spam Collection is a set of SMS tagged messages that have been collected for SMS Spam research. It contains one set of SMS messages in English of 5,574 messages, tagged as being ham (legitimate) or spam.

Training and Parameters

Batch size = 16

Epochs = 3

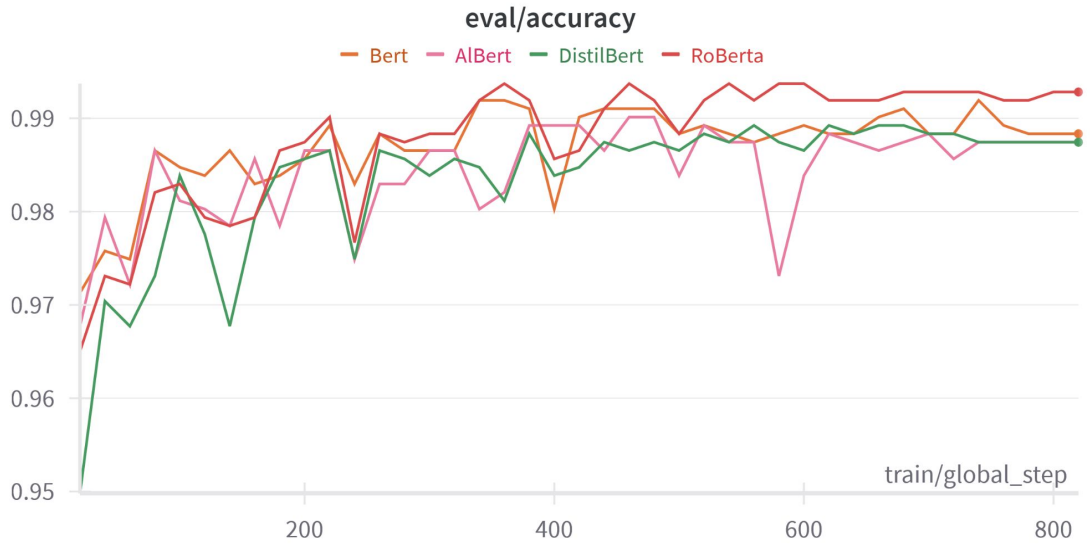
Weight decay = 0.01

Learning Rate = 2e-5

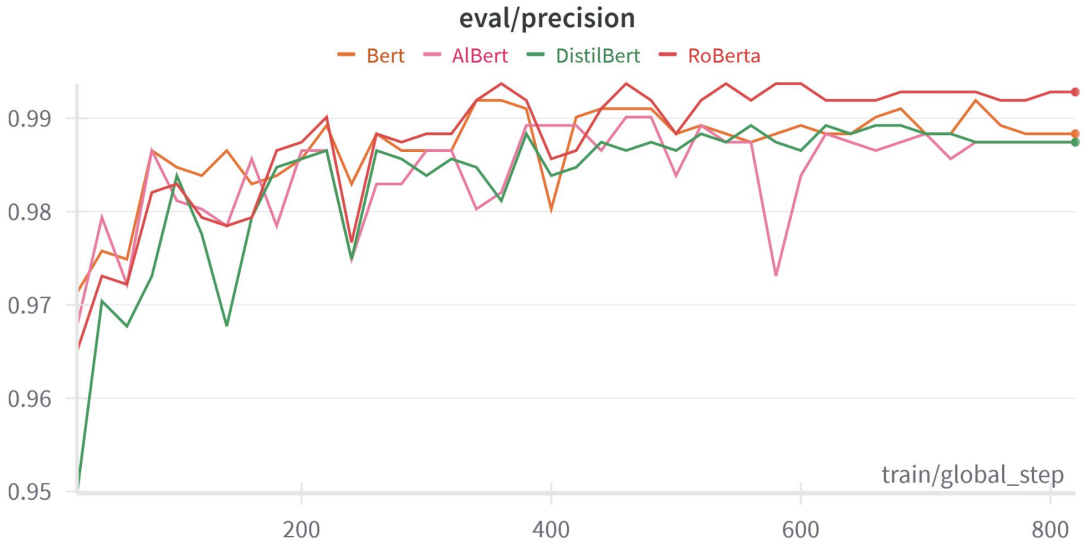
Evaluation Steps = 20

Save Steps = 800

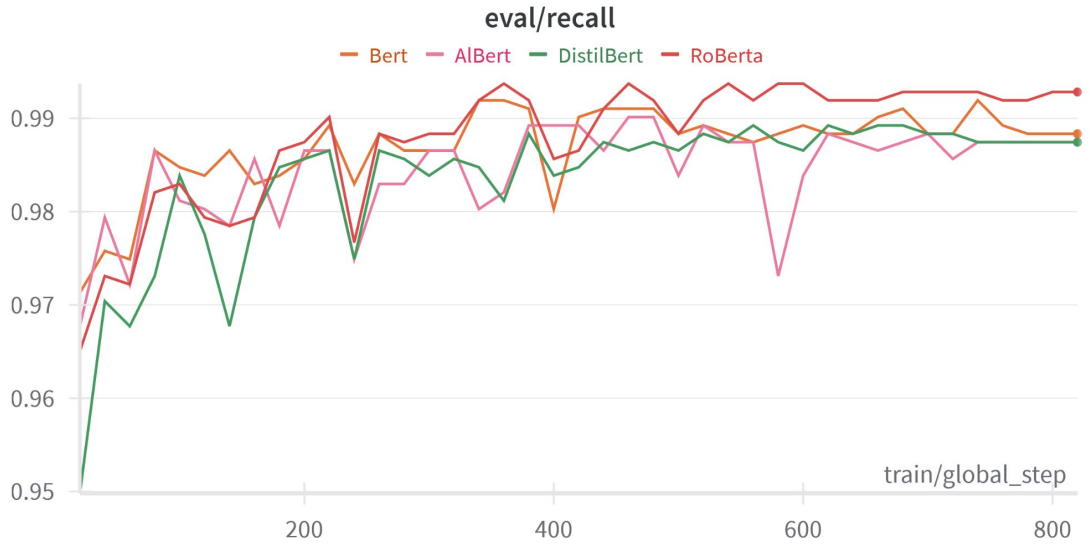
Results


eval/accuracy
Bert
AlBert
DistilBert
RoBerta
0.99
0.98
0.97
0.96
0.95
train/global_step
200
400
600
800

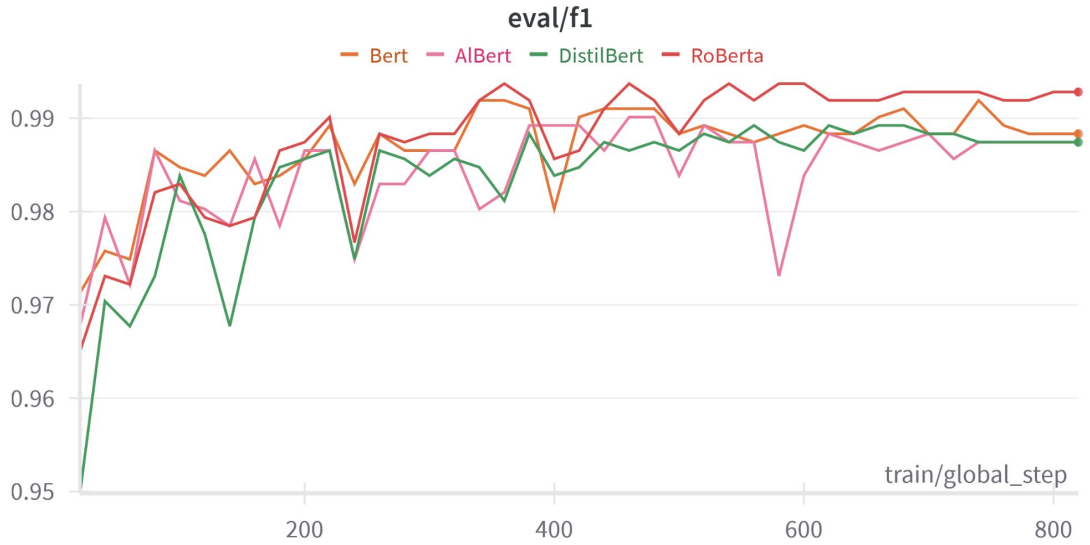
Results


eval/precision
Bert
AlBert
DistilBert
RoBerta
0.99
0.98
0.97
0.96
0.95
train/global_step
200
400
600
800

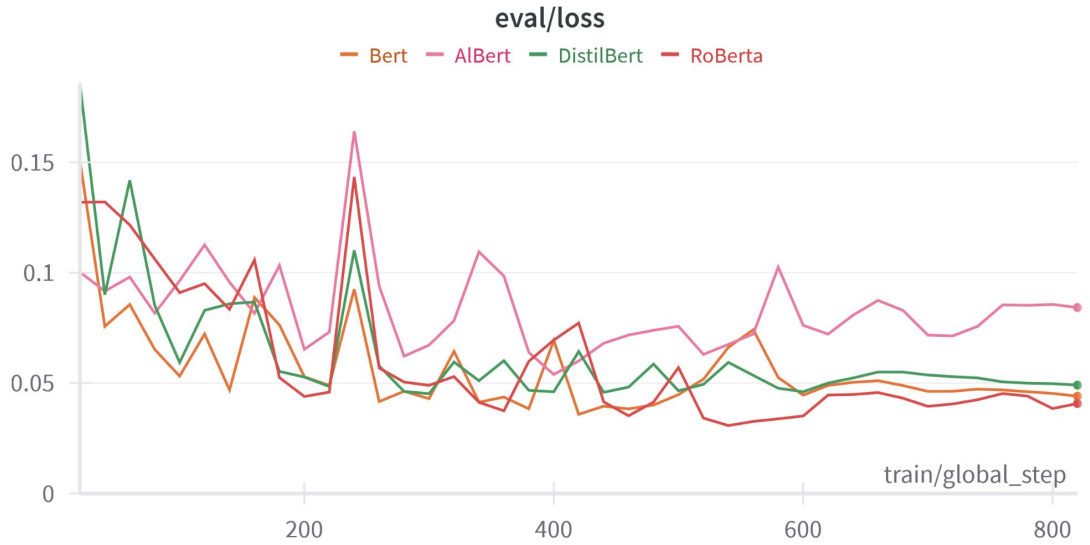
Results


eval/recall
Bert
AlBert
DistilBert
RoBerta
0.99
0.98
0.97
0.96
0.95
train/global_step
200
400
600
800

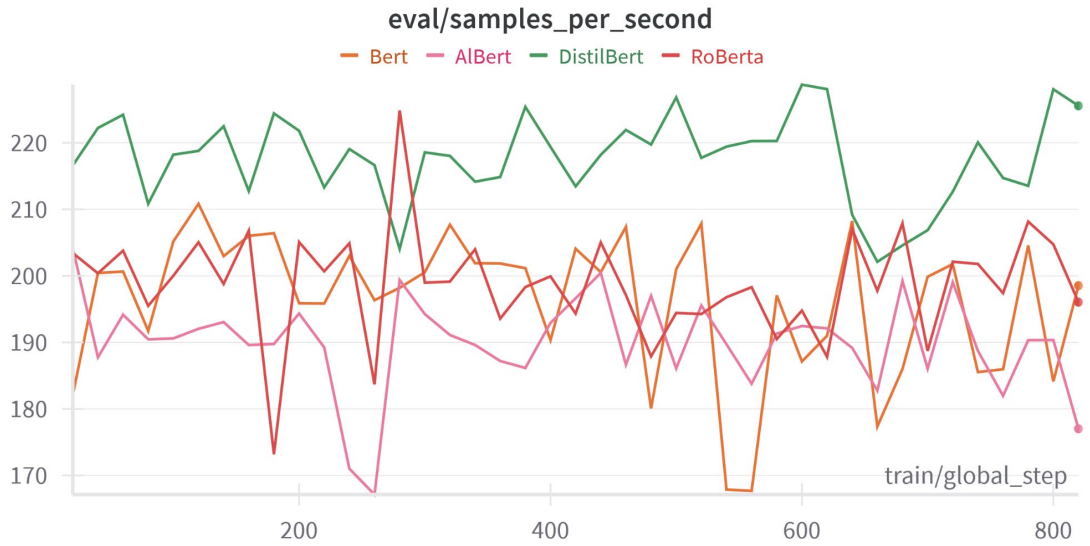
Results


eval/f1
Bert
AlBert
DistilBert
RoBerta
0.99
0.98
0.97
0.96
0.95
train/global_step
200
400
600
800

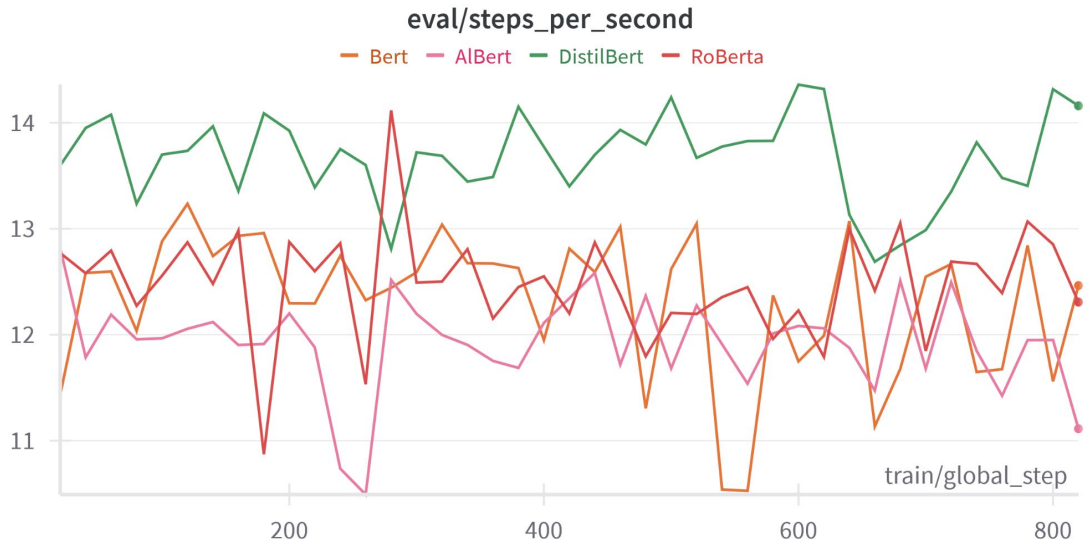
Results


eval/loss
Bert
AlBert
DistilBert
RoBerta
0.15
0.1
0.05
0
train/global_step
200
400
600
800

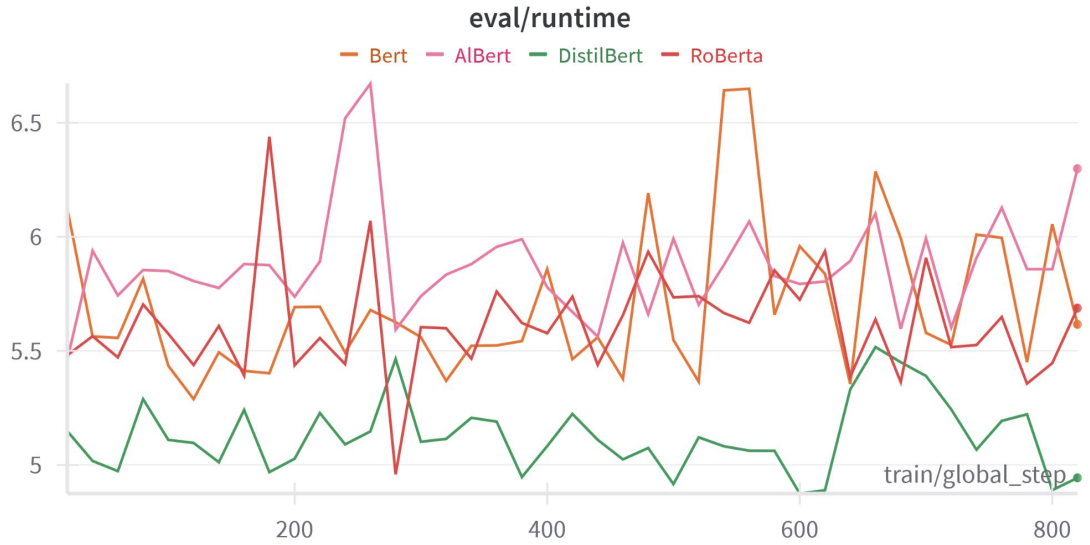
Results


eval/samples_per_second
Bert
AlBert
DistilBert
RoBerta
220
210
200
190
180
170
train/global_step
200
400
600
800

Results


eval/steps_per_second
Bert
AlBert
DistilBert
RoBerta
14
13
12
11
train/global_step
200
400
600
800

Results


eval/runtime
Bert
AlBert
DistilBert
RoBerta
6.5
6
5.5
5
200
400
600
800
train/global_step