Results of fine tuning on CNN\_DailyMail:

max\_input\_length = 1024

max\_target\_length = 128

google/t5-v1\_1-small, 20k - train, epochs - 2

| Grouping | Num KV heads | Val loss | R1 score - test | R1 score - val |
| --- | --- | --- | --- | --- |
| Standard | 3 | 2.438816 | 21.64 | 21.691900 |
| Close attention | 3 | 2.454108 |  | 21.722400 |
| Far attention | 3 | 2.443279 | 21.63 | 21.591 |

Results of fine tuning on XSum:

| Model | Num KV heads | Dataset size | Epochs | R1 score |
| --- | --- | --- | --- | --- |
| t5-small | MHA | 50k - train | 3 |  |
| t5-small | 4 | 50k - train | 3 | 29.39 |
| google/t5-v1\_1-small | MHA | 20k - train | 1 | 26.06 |
| google/t5-v1\_1-small | 3 | 20k - train | 1 | 23.98 |
| google/t5-v1\_1-small | MHA | 30k - train | 2 | 26.96 |
| google/t5-v1\_1-small | 3 | 30k - train | 2 | 25.72 |