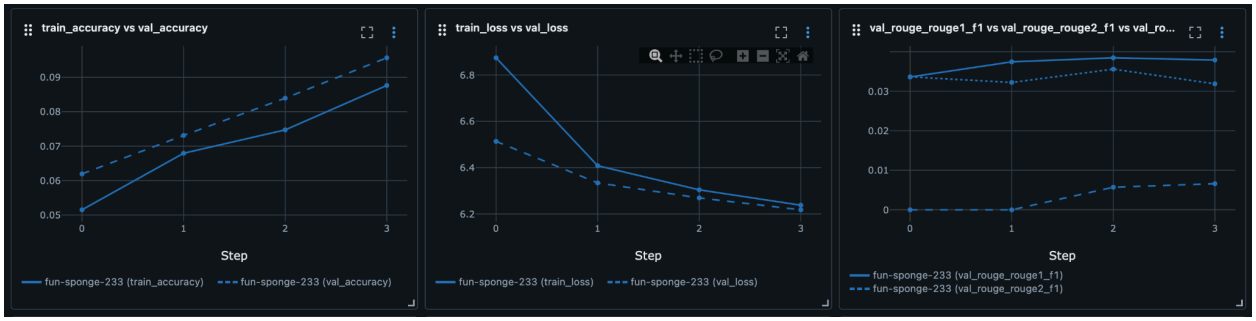
# Encoder-only Model Appendix

## Additional Decoder-only Experiments

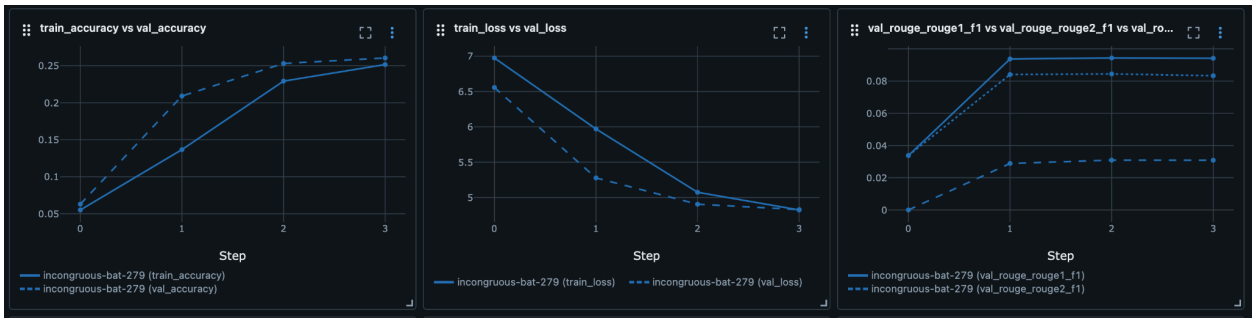
### Vocab Size

We run experiments with three different vocab sizes (8k, 12k, 15k) keeping other parameters equal. Here are the graphs and scores for the three experiments.

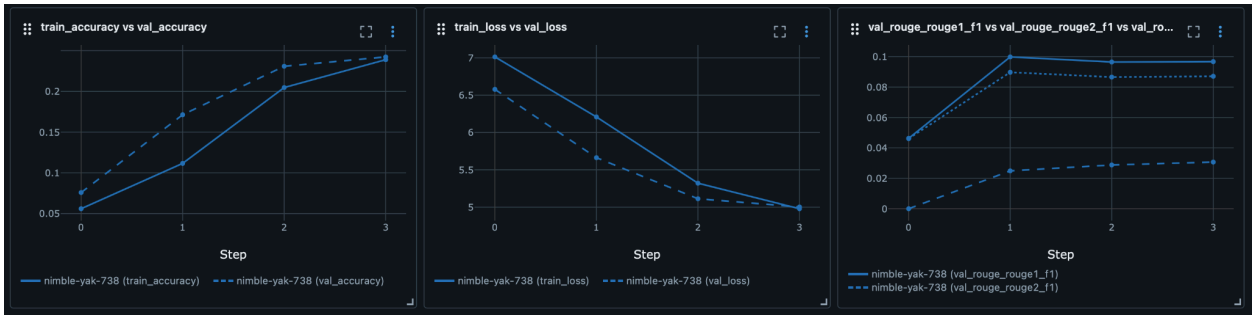
fun-sponge-233 - 8k - train_loss=6.23 val_loss=6.12 val_acc= 0.09, val_rougel_f1=0.03




Incongruous-bat-279 - 12k - train_loss=4.82 val_loss=4.82 val_acc=0.26 val_rougel_f1=0.083




Nimble-yak-738 - 15k - train_loss=4.97 val_loss=5.00, val_acc= 0.24, val_rougel_f1=0.087

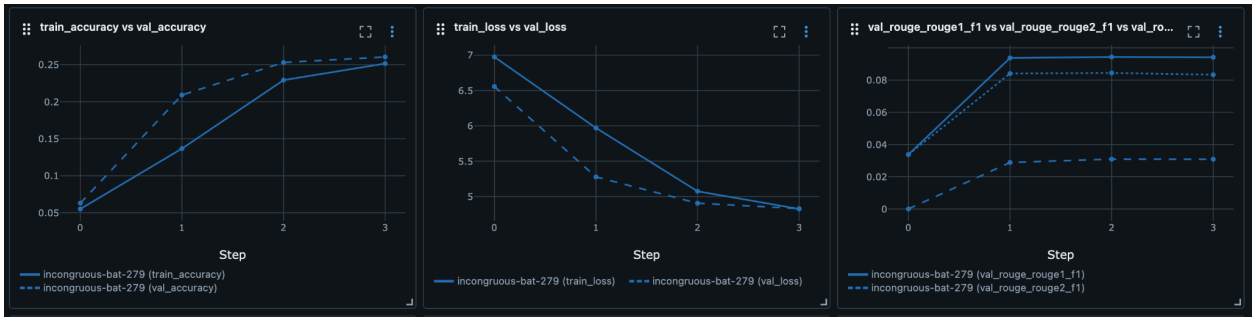


As you can see, even though SentencePiece says that all three options have a character coverage over 99%, in practice increasing the vocabulary size improves validation accuracy, loss and rouge scores. However, given our compute restrictions and the small delta between 12k and 15k tokens, we use the 12k models as baselines for hyperparameter analysis.

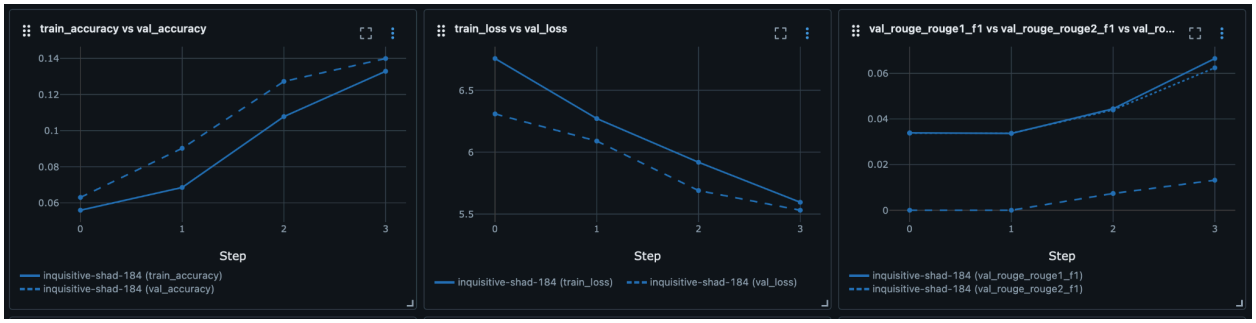
## Regularisation

Though we apply a relatively soft regularisation on our base models, it's interesting to see the effect of no regularisation at all on performance.

Incongruous-bat-279 - Base Regularisation - train_loss=4.82 val_loss=4.82 val_acc=0.26 val_rougel_f1=0.08




inquisitive-shad-184 - No Regularisation (dropout, weight_decay, label_smoothing = 0) - train_loss=5.59 val_loss=5.53 val_acc=0.13 val_rougel_f1=0.06



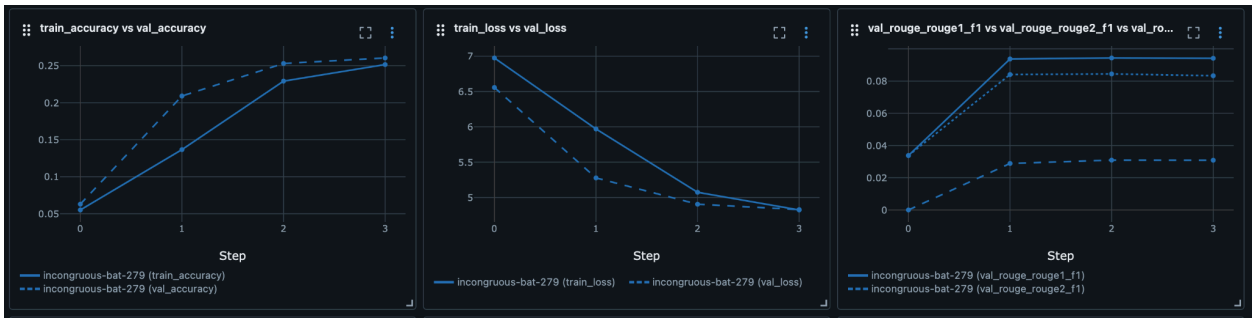

After not regularisation performance is still pretty similar, with a small drop in validation metrics. In practice, the small capacity helps the model avoid overfitting to training examples, which is clear in the train-loss to eval-loss ratio.

## Gradient Accumulation

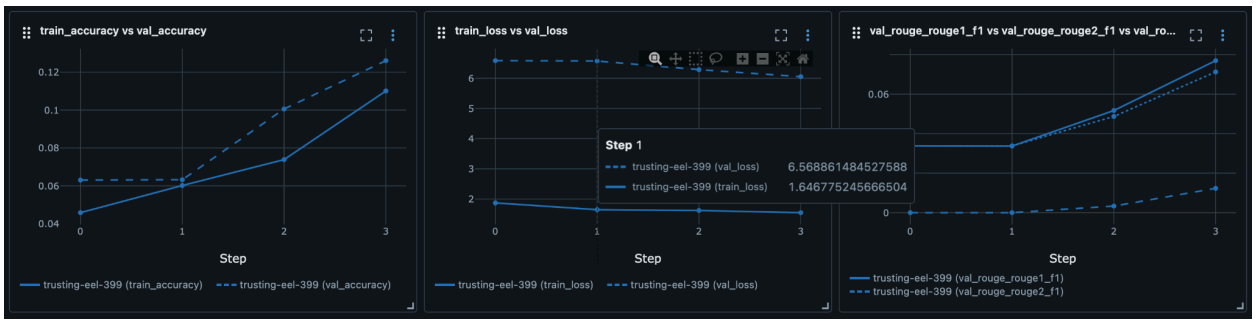
Some researchers recommend having a larger batch size when training transformer models, especially with large amounts of data, to reduce the variations between gradients and converge

with fewer jumps. Since our compute resources are limited, we use the gradient accumulation trick to understand the impact of larger updates for our default configuration.

Incongruous-bat-279 - No Accumulation - train_loss=4.82 val_loss=4.82 val_acc=0.26 val_rougel_f1=0.08




trusting-eel-339 - Accumulate every 4 minibatches - train_loss=1.54 val_loss=6.05 val_acc=0.12 val_rougel_f1=0.07




Though training loss is significantly reduced, validation metrics perform worse, despite multiple additional experiments. We don't see significant improvement in visual tests either.

## Beam Search

Beam search is a common method to improve sequence decoding performance. However, it's best suited when the model already generates good outputs. Since we are not there yet we didn't expect significant improvements, and we were correct.

Base Model Summary - Beam 1:

- Existing law requires the state, and the state. This bill would require the state. The bill would require the state-mandated local program…
- Existing law requires the state. The California Constitution requires the state. This bill would require the state-mandated local agencies…

Base Model Summary - Beam 4:

- This bill would bill would bill would bill would bill would require the bill would bill…
- This bill would bill would bill would bill would bill would bill would bill would bill…

It is clear that instead of improving the outcomes moves further into the local minima of predicting the most common tokens in the summaries dataset.

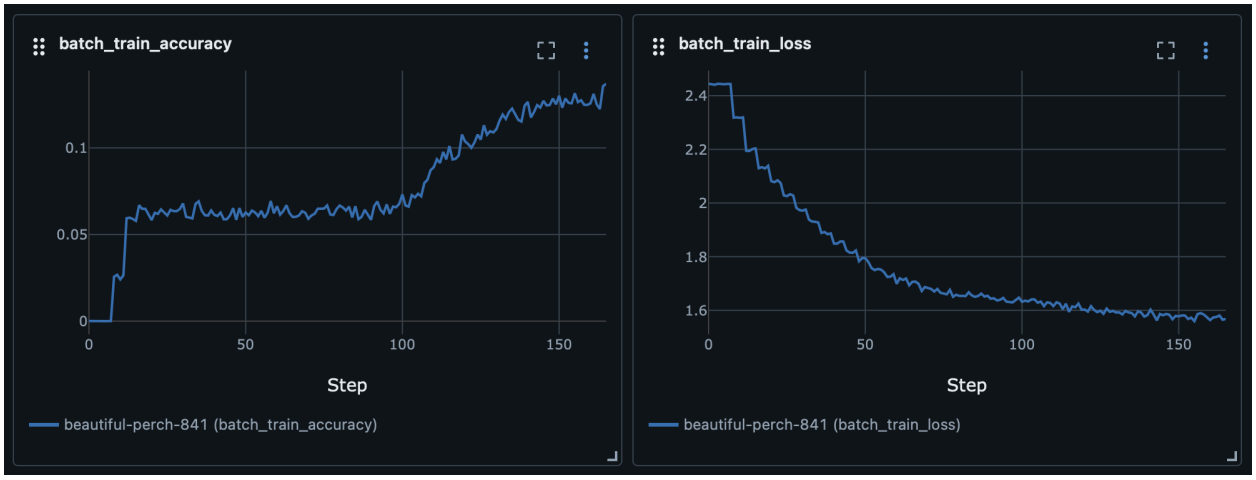
## Increasing Model Capacity

We tried to increase model capacity in multiple ways. The first experiment where we only increased the number of layers and dimensions failed as expected, since a larger model has implications on regularisation, batch size and other hyperparameters.

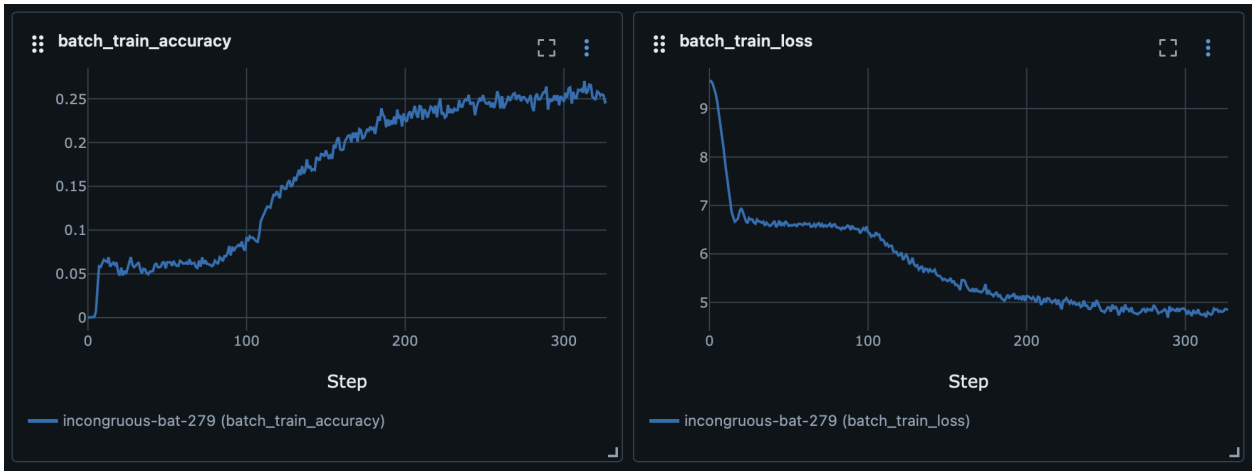
After significant search, here's the best larger model we could train with limited compute and data, which is unable to beat our base model. Since we were only able to train for two epochs, here are the hyperparameters, batch graphs and final results.

```
clean_diff_text: True
vocab_size: 15000
dim_size: 256
dim_feedforward: 1024
num_layers: 4
num_heads: 4
dropout: 0.1
learning_rate: 0.001
weight_decay: 0.05
warmup_steps: null
max_len: 1024
min_summary_length: 64
epochs: 2
batch_size: 64
grad_accum_steps: 4
label_smoothing: 0.05
beam_size_val: 1
beam_size_sample: 1
length_penalty: 0.6
tokenizer: sentencepiece
```

beautiful-perch-841 - Large Model - train_loss=1.60 val_loss=6.25 val_acc=0.13 val_rougel_f1=0.07




Incongruous-bat-279 - Base model - train_loss=4.82 val_loss=4.82 val_acc=0.26 val_rougel_f1=0.08




As it's clear in the results, gradient accumulation keeps training loss low but validation loss high. However, when we try to increase regularisation to improve validation performance, the model stops learning useful outputs, even if the train loss keeps being low.

## Best Decoder Sample outputs

**Example Diff:** SECTION 1. Section 1182.12 of the Labor Code is amended to read:

1182.12. <ins_start> (a) <ins_end> Notwithstanding any other provision of this part, on and after January 1, <del_start> 2007 <del_end> <ins_start> 2014 <ins_end> , the minimum wage for all industries shall be not less than <del_start> seven dollars and fifty cents ($7.50) per hour, and on and after January 1, 2008, the minimum wage for all industries shall be not less than eight dollars ($8.00) per hour <del_end> <ins_start> eight dollars and twenty-five cents ($8.25) per hour and on and after January 1, 2015, the minimum wage for all industries shall be not less than eight dollars and seventy-five cents ($8.75) per hour and on and after January 1, 2016, the minimum wage for all industries shall be not less than nine dollars and twenty-five cents ($9.25) per hour <ins_end> .

<ins_start> (b) (1) Except as provided in paragraph (3), the minimum wage shall be automatically adjusted on January 1 of each year, commencing on January 1, 2017, to maintain employee purchasing power diminished by the rate of inflation that occurred during the previous year.

(2) The minimum wage adjustment shall be made by multiplying the minimum wage in effect on December 31 of the previous year by the percentage of inflation that occurred during that year, and by adding the product to the wage in effect during that year. The resulting total shall be rounded off to the nearest five cents ($0.05).

The Industrial Welfare Commission shall publicize the automatically adjusted minimum wage.

(3) The Industrial Welfare Commission shall not adjust the minimum wage pursuant to this subdivision if the average percentage of inflation for the previous year was negative.

(4) For purposes of this subdivision, the following terms have the following meanings:

(A) “Percentage of inflation” means the percentage of inflation specified in the California Consumer Price Index for All Urban Consumers, as published by the Department of Industrial Relations, Division of Labor Statistics and Research, or its successor index.

(B) “Previous year” means the 12-month period that ends on August 31 of the calendar year prior to the adjustment.

(c) The Industrial Welfare Commission shall not reduce the minimum wage prescribed by this section.

(d) This section shall not be construed to preclude an increase of the minimum wage by the Industrial Welfare Commission in an amount that is greater than the rate calculated pursuant to subdivision (b). <ins_end>

**Example Summary:** Existing law requires the state, and the state. This bill would require the state. The bill would require the state-mandated local program. This bill would require the state-mandated local agencies and school districts for the state. This bill would require the state. The bill would require the state, and would require the state-mandated local agencies and requires the state. The bill would require the state-mandated local agencies and would require the state. This bill would require the state. The bill would require the state. This bill would require the state. This bill would require the state. The bill would require the state. The bill would require the state. The bill would require the state-mandated local program. This bill would require the state. This bill would require the state. The bill would require the state to the state. This bill would require the state to the state-mandated local agencies and the state to the state, and the state. The bill would require the state. This bill would require the state to the state. This bill would require the state. This bill would require the state. The bill would require the state. The bill would require the state board to the state to the state. This bill would require the state to the state-mandated local program. The bill would require the state. The bill would require the state. The bill would require the state. The bill would require the state to the state-mandated local program. The bill would require the state-mandated local program. The bill would require the state. The bill would require the state-mandated local agencies and the state-mandated local agencies and the state to the state-mandated local agencies and the state-mandated local program. This bill would require the state. The bill would requir

**Example Diff:** SECTION 1. Section 9250.20 is added to the Vehicle Code, to read:
9250.20. (a) In addition to any other fees specified in this code and the Revenue and Taxation Code, a fee of six dollars ($6) shall be paid at the time of registration or renewal of registration of every vehicle subject to registration under this code, except vehicles described in subdivision (a) of Section 5014.1 and those vehicles that are expressly exempted from the payment of registration fees under this code.
(b) After deducting all costs incurred pursuant to this section, the department shall remit all money realized pursuant to this section for deposit in the Sustainable Communities Strategy Subaccount which is hereby established in the Motor Vehicle Account and made available, upon appropriation by the Legislature, for implementation of sustainable communities strategies adopted pursuant to paragraph (2) of subdivision (b) of Section 65080 of the Government Code.
SEC. 2. This act is an urgency statute necessary for the immediate preservation of the public peace, health, or safety within the meaning of Article IV of the Constitution and shall go into immediate effect. The facts constituting the necessity are:
In order to provide immediate funds for agencies and communities seeking to implement sustainable communities strategies, it is necessary that this act take effect immediately.
**Example Summary:** Existing law requires the state. The California Constitution requires the state. This bill would require the state-mandated local agencies and the state-mandated local agencies and the state, and the state. The bill would require the state-mandated local program. The bill would require the state-mandated local agencies and the state. The bill would require the state, and the state board to the state-mandated local agencies and the state, and the state. The bill would require the state-mandated local program. This bill would require the state-mandated local agencies and the state, and the state-mandated local agencies and the state. This bill would require the state. This bill would require the state-mandated local program. The bill would require the state-mandated local agencies and the state-mandated local program. The bill would require the state agency to the state. The bill would require the state. The bill would require the state-mandated local agencies and the state. The bill would require the state-mandated local program. This bill would require the state-mandated local program. The bill would require the state-mandated local agencies and the state-mandated local agencies and the state-mandated local program. The bill would require the state. The bill would require the state. The bill would require the state. The bill would require the state. The bill would require the state. The bill would require the state-mandated local program. The bill would require the state-mandated local agencies and the state. This bill would require the state-mandated local program. The bill would require the state. The bill would require the state-mandated local program. The bill would require the state-mandated local program. The bill would require the state-mandated local agencies and the state-mandated local program. This bill would require the state-mandated local agencies and the state. The bill would require the state. This bill would require the state. This bill would require the state, and the state board to the state-mandated local agencies and the state. The bill would require the state-mandated local agencies and the state. The bill would require the state. This bill would require the state to the state. The bill would require the state board to the state-mandated local agencies and the state to the state to the state-mandated local agencies and the state to the state to the state. Existing law requires the state-mandated local agencies and the state. This bill would require the state. This bill would require the state-mandated local agencies and the state-mandated local agencies and the state. The bill would require the state-mandated local agencies and the state. The bill would require the state-mandated local agencies and would require the state. The bill would require the state agency to the state to the state. Existing law requires the state to the state-mandated local agencies and would require the state. Existing law requires the state-mandated local program. The bill would require the state. The bill would require

the state-mandated local agencies and the state-mandated local agencies to the state to the state-mandated local agencies and the state-mandated local program. This bill would require the state. The bill would require the state to the state-mandated local agencies and the state-mandated local agencies and the state to the state. The bill would require the state. The bill would require the state. The bill would require the state to the state to the state. The bill would require the state. The bill would require the state to the state, and the state-mandated local agencies and the state. This bill would require the state to the state. The bill would require the state. The bill would require the state. The bill would require the state. The bill would require the state-mandated local program. The bill would require the state