

Review 1

We would like to thank the reviewer for their efforts in improving our manuscript. We have edited the text in response to the reviewer's comments as follows.

It is not clear whether the authors have investigated all the variations of Transformer.

When considering the scope of the paper, which is pedagogical and focused on process transparency, the short answer is yes. This is because we adopted the architecture proposed by Vaswani et al. (arXiv, 2017), which lies at the heart of all Transformer architectures; all others add components to this basic framework. However, in response to the reviewer's request for focus, we have narrowed the declared scope of our contribution, beginning with the title and abstract, and revising several points in the text. These and all other changes are highlighted in blue.

It is not clear what is the fundamental novelty of the proposal.

This contribution does not represent another improvement to the basic framework, but rather a better understanding of the core mechanisms of fundamental processes within the core structure. To the best of our knowledge, transformers have not previously been described in complete pseudocode, nor have their memory requirements been detailed at the level of each individual learnt parameter. Furthermore, even the simplest vanilla time series forecasting architectures available for download include some optimized details, whereas our code is simply the most basic implementation of the architecture proposed in Vaswani et al. (arXiv, 2017) when applied to time series forecasting. We believe that our contribution is valuable to the research community because it provides an unquestionable baseline against which to validate the results of more elaborate additions, all the more so since even such a basic implementation generates results of nontrivial quality. We have tried to make these points clearer in the edited manuscript.

It is not clear whether the authors have made sufficient comparisons with all the variations of Transformer.

In addition to the points outlined above, we now also provide a computational comparison of the results obtained using our minimalist code and those obtained using two advanced time series forecasting architectures. While these are not 'all the variations', which currently number in the dozens with more being added, they nevertheless enable us to evaluate the difference in quality between the results obtained using our minimalist code and those at the forefront of the field.