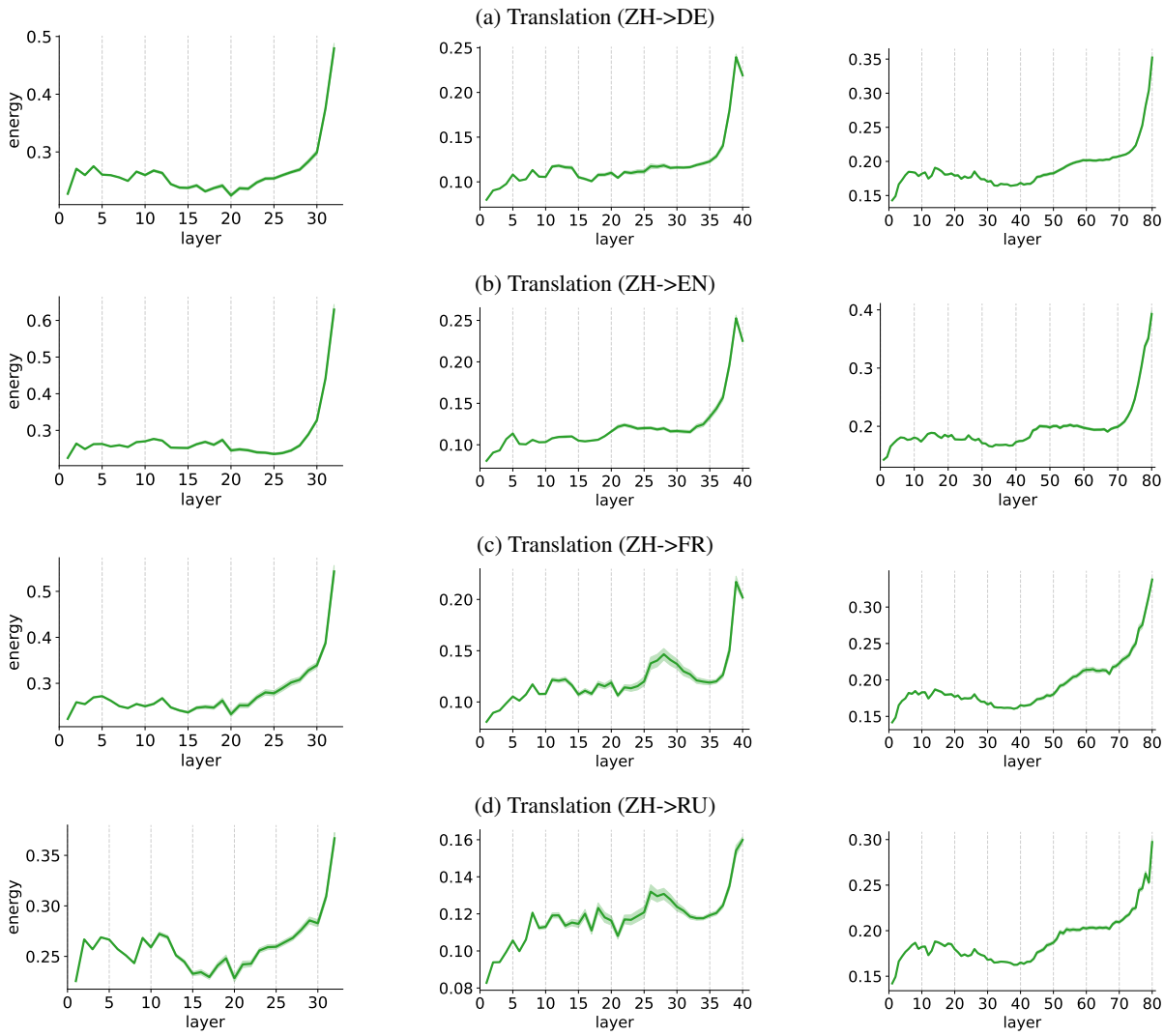

Figure 20: Figures illustrate the translation task where Llama-2 7B, 13B, and 70B are tasked with translating a word from non-English input language to output language. There is one column per model size. The x-axis shows the layer number of the model, and the y-axis the energy. Means and 95% Gaussian confidence intervals have been computed over the input examples, numbers in Appendix A.

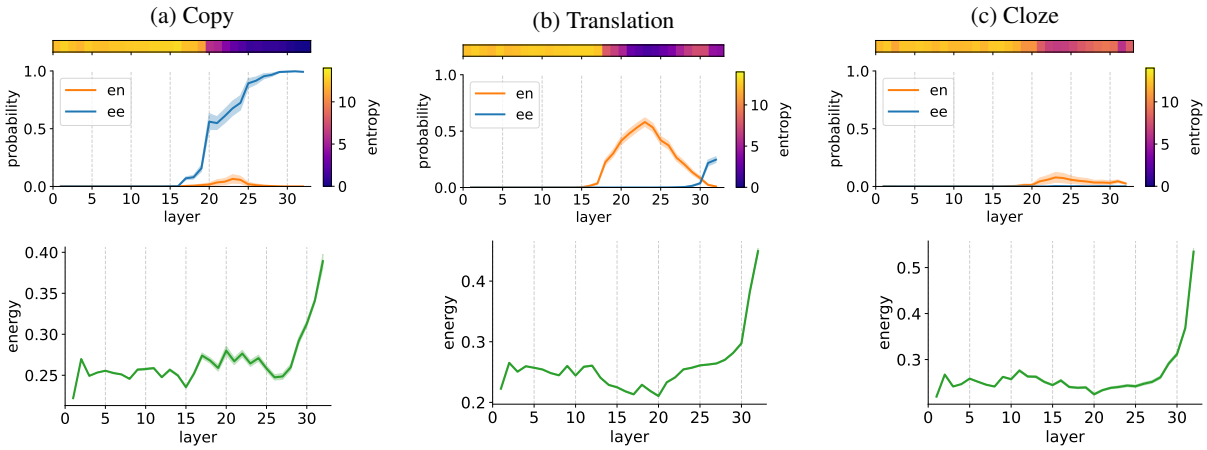

Figure 21: Figures illustrate our analysis of the copy-, translation-, and cloze task for the **Estonian** language on Llama-2-7B. In the first row, the x-axis shows the layer number of the model, and the y-axis the language probability. In the first row, the x-axis shows the layer number of the model, and the y-axis the token energy. Means and 95% Gaussian confidence intervals have been computed over the input examples.

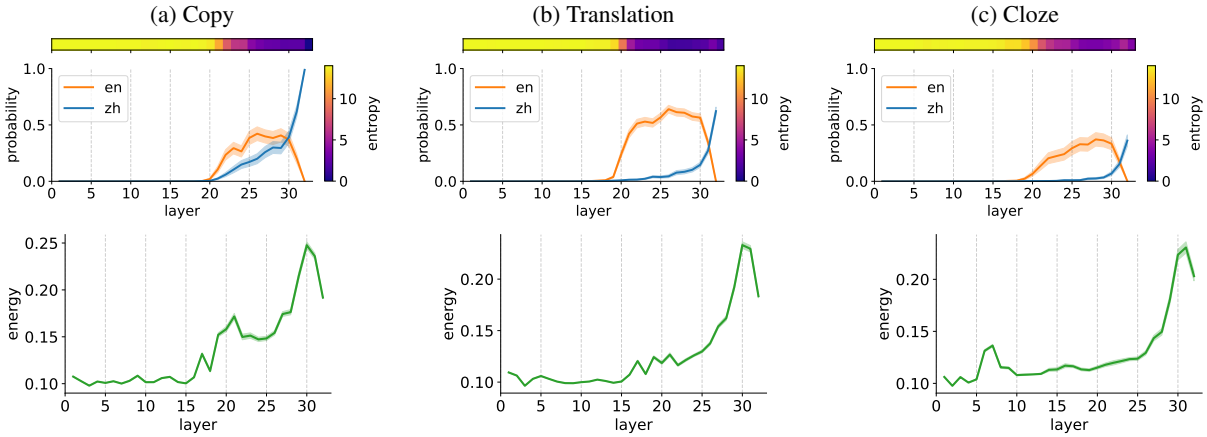

Figure 22: Figures illustrate our analysis of the copy-, translation-, and cloze task for Chinese on **Mistral-7B**. In the first row, the x-axis shows the layer number of the model, and the y-axis the language probability. In the first row, the x-axis shows the layer number of the model, and the y-axis the token energy. Means and 95% Gaussian confidence intervals have been computed over the input examples.