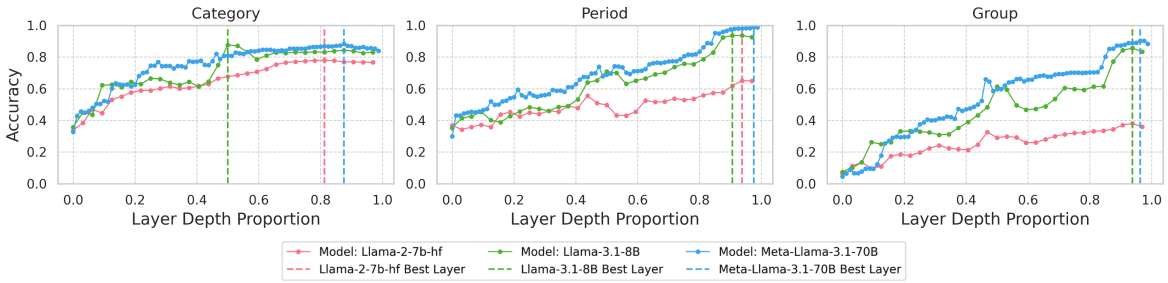

Figure 1: R^2 score trends. Categorical linear probing on the last token activation with 5-fold cross-validation was performed on activations, and R2 scores on the test set are shown for each attribute.

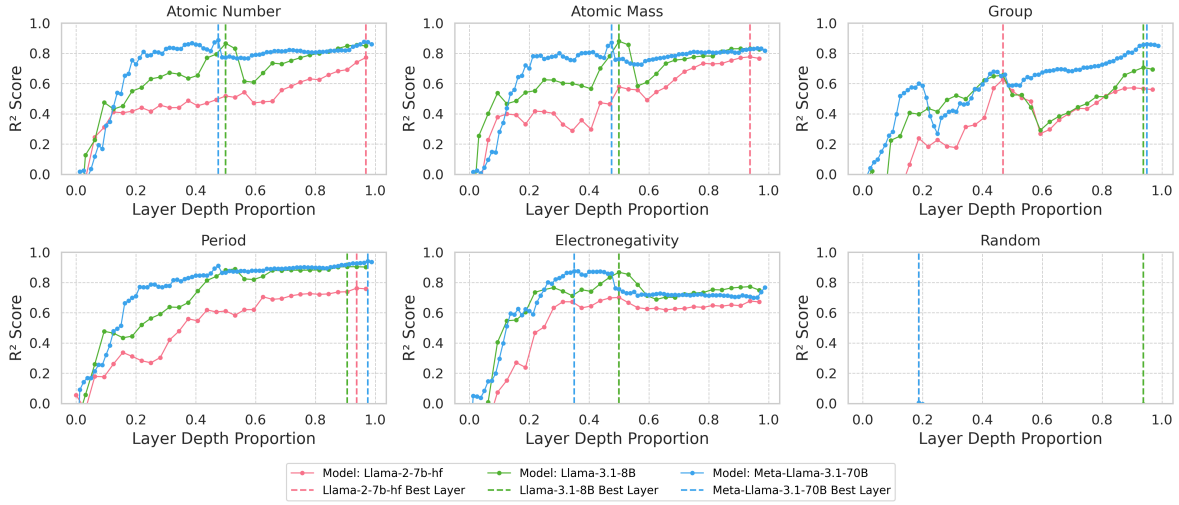

Figure 2: R^2 score trends. Regression linear probing on the last token activation with 5-fold cross-validation was performed on activations, and R2 scores on the test set are shown for each attribute.

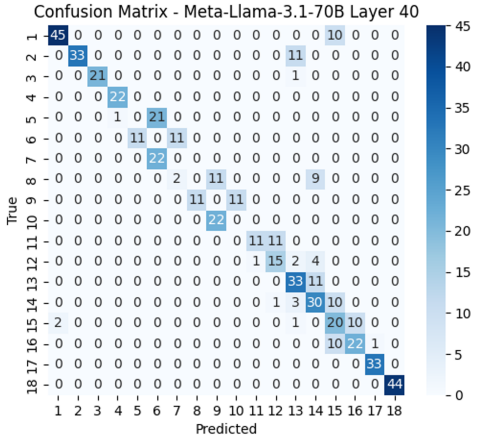

Figure 3: Confusion Matrix on categorical linear probing of attribute 'Group' on the middle layer.

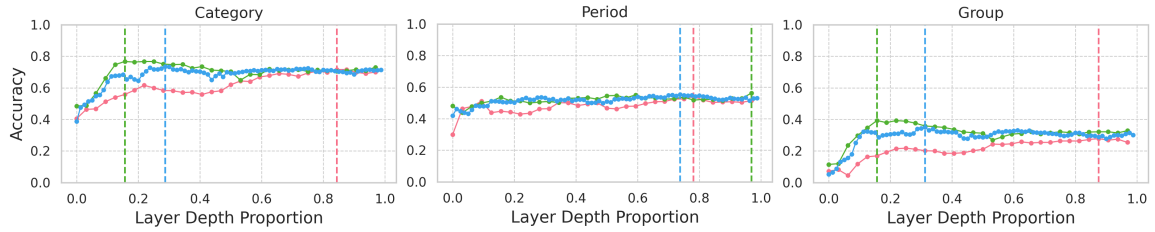

Figure 4: R^2 score trends. Categorical linear probing on the element token activation with 5-fold cross-validation was performed on activations, and R2 scores on the test set are shown for each attribute.

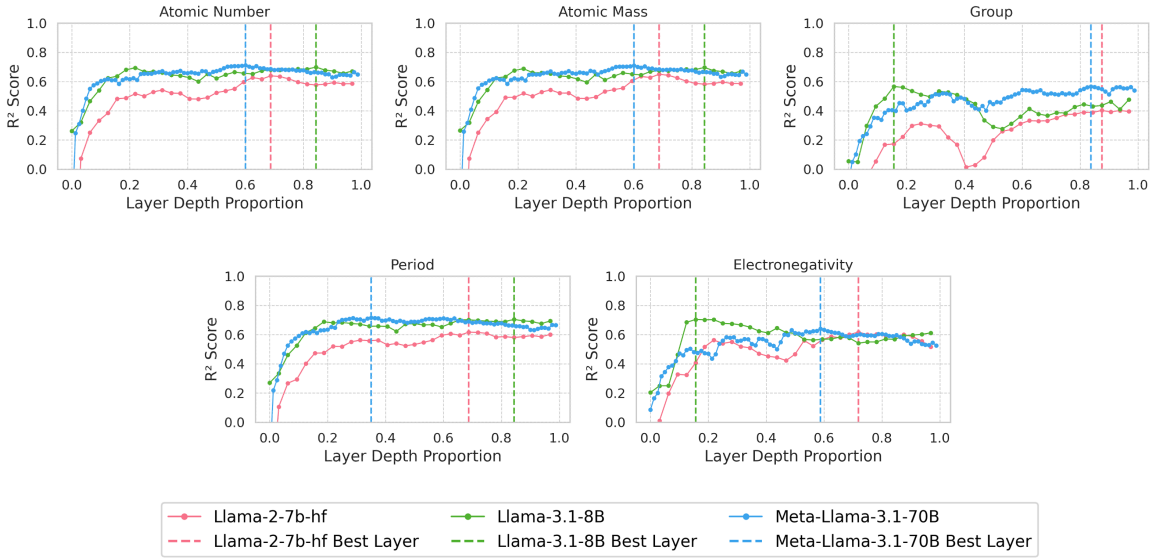

Figure 5: R^2 score trends. Regression linear probing on the element token activation with 5-fold cross-validation was performed on activations, and R2 scores on the test set are shown for each attribute.