

ABSTRACT

The present application provides a data processing method and apparatus based on a large language model, and a device. The method comprises: obtaining an original input question, and determining a target feature corresponding to the original input question; selecting a target branch model corresponding to the target feature from among all branch models; inputting the original input question to a basic model, so as to obtain an intermediate processing feature corresponding to the original input question and an output processing feature corresponding to the original input question; and sending a trigger instruction to the target branch model, so that the target branch model obtains the intermediate processing feature and the output processing feature from the basic model on the basis of the trigger instruction, processes the intermediate processing feature so as to obtain a branch processing feature corresponding to the original input question, and processes the branch processing feature and the output processing feature so as to obtain a target answer corresponding to the original input question. The technical solution of the present application saves computing resources of a server and time overhead, and thus a target answer can be quickly determined.

[Fig. 1]