Mask Rate = 90%

Original

After the release of ChatGPT in 2022, the number of papers published every day at pout Large Language Models (LLMs) has in creased more than 20-fold. The number of of parameters in these LLMs jumped from 340 millions in implementations such as B ERT to billions of parameters in models like GPT-3 or LLaMA. A large part of these page arameters come from the word-embeddin ig layers which are used to represent a fin ite vocabulary of characters, sets of chara cters or words. Apart from increasing mod del complexity, a fixed vocabulary is also re esponsible for brittle models, which cannot t deal with out-of-vocabulary inputs and ca nnot generalize to new languages. As a

Original + SD

After the release of ChatGP number of papers published every day a ut Large Language Model ore tha 0.0-fold. The heters in these LLMs jumped from 340 millions in implementations such as B ERT to billions of paramete in models lil PT-3 or LLaMA. A large part of these i from the word-embeddi arameters ig layers which are used to represent ite vocabular of corracter sets of chara cters or words. Apart from increasing mo mplexity, a fixed vocable ry th out-of-vocab

Mean Prediction + SD

