Now anyone can evaluate and test large language models (LLMs) on their own.

Upstage (CEO Kim Sung-hoon) announced on the 3rd that it will release its no-code LLM comprehensive evaluation platform 'Evalverse' for free. It is released in the form of open source and is available on GitHub, so anyone can easily use it.

Usually, LLM performance evaluation uses 'benchmark' indicators that record response accuracy based on specific datasets. Major benchmarks include Hugging Face's 'H6' indicator, which measures six categories such as the model's reasoning, common sense, and language understanding ability, as well as the conversation ability indicator 'MT-bench', the emotional evaluation indicator 'EQ-bench', and the instruction fulfillment ability indicator 'IFEval'.

In response, Upstage has developed its own evaluation platform that integrates these major benchmarks to comprehensively test language model performance. Previously, it was cumbersome to submit models to the corresponding platform to test specific benchmarks, but with Evalverse, multiple benchmark tests can be conducted simultaneously.

Upstage Evalverse has expanded its usability by linking it to the API chatbot in the business messenger 'Slack'. In particular, it is characterized by simplifying all evaluation processes through 1:1 conversations (DM) with the Evalverse chatbot within Slack. Users can call and test the desired model by entering only two commands, 'call' and 'evaluate', in the chat window, and can also check a comprehensive report on the score for each benchmark, making it easy to compare and evaluate between models.

Upstage's release of Evalverse as open source is part of its efforts to develop and coexist with the generative AI ecosystem. Upstage has completed its unique open source worldview by releasing the evaluation platform after building the 'Open Ko-LLM Leaderboard', which presents the standard for

evaluating Korean LLMs, and 'Dataverse', which fully discloses complex data preprocessing technologies.

Upstage CEO Kim Sung-hoon said, "We are pleased to release the Evalverse platform, which allows anyone to easily test the LLMs they have developed, for free." He added, "In the future, Upstage will continue to make various efforts to develop and spread the LLM ecosystem as a leader in the global generative AI market."