

The 'Open Ko-LLM Leaderboard', jointly operated by Upstage (CEO Kim Sung-hoon) and the National Information Society Agency (NIA, Director Hwang Jong-sung), will be fully revamped into Season 2 with new benchmarks from August 12th.

The 'Open Ko-LLM Leaderboard' is the largest open Korean language model (LLM) evaluation system in Korea, which was opened in September 2023 through public-private cooperation, and interest and participation are continuing with more than 1,700 LLM models submitted from various fields of industry, academia, and research as of the end of July.

Recently, as LLM technology has developed rapidly, the need for improvement of the leaderboard benchmark created a year ago has emerged, and as a result, Upstage and NIA have decided to completely abolish the existing benchmark and start operating the Open Ko-LLM Leaderboard Season 2 with new performance evaluation indicators.

The newly added benchmarks are a total of nine, including ?Ko-GPQA (graduate-level reasoning) ?Ko-WinoGrande (common sense reasoning) ?Ko-GSM8K (elementary school math) ?Ko-EQ-Bench (emotion) ?Ko-IFEval (instruction fulfillment) ?KorNAT-Knowledge (Korean reliability) ?KorNAT-Social-Value (social value) ?Ko-Harmlessness (harmlessness) ?Ko-Helpfulness (practicality).

While the previous version focused on evaluating the fundamental language ability of LLM, the new season focuses on practical and realistic aspects such as more advanced intelligence and problem-solving abilities. It is expected that it will be able to distinguish between high-performance models by composing indicators for this purpose.

Accordingly, all LLMs submitted in the future, including all models submitted in the past, will be evaluated through the new benchmark. In addition, AI data specialist companies Flitto, Selectstar, and KAIST AI

Graduate School will participate in the new season and provide specialized benchmark datasets.

Upstage CEO Kim Sung-hoon said, "We expect the newly revamped leaderboard to raise the benchmark for evaluating the performance of Korean LLMs to the next level." He added, "In the future, Upstage will take the lead in strengthening the domestic AI ecosystem that surpasses global standards."

NIA Director Hwang Jong-sung said, "The leaderboard has been greatly contributing to the creation of the Korean LLM ecosystem through public-private collaboration, and we hope that the season 2 revamp will be an opportunity for Korean LLM technology to take a step further to the global level."