

Upstage and the National Information Society Agency (NIA) are joining hands to create a leaderboard that can evaluate and compare the performance of Korean LLMs to promote the competitiveness of Korean AI.

Upstage (CEO Kim Sung-hoon) announced that it will establish and open the Korean LLM leaderboard 'Open Ko-LLM Leaderboard' on September 27, co-hosted by NIA. This is the result of the cooperation between Upstage and NIA on the 4th, and Upstage plans to continue to expand the Korean AI ecosystem by adding the open LLM leaderboard to the 1T club.

The Open Ko-LLM Leaderboard is an open platform where anyone can register and compete with their own Korean LLM models. Researchers interested in the Open Ko-LLM Leaderboard can check the details and participate in the Open Ko-LLM Leaderboard space on Huggingface after the 27th.

Upstage and NIA's Open Ko-LLM Leaderboard has the advantage of being a Korean-specific leaderboard by building high-quality data that reflects the characteristics and culture of Korean language, not just translating the existing data of the open LLM leaderboard operated by Huggingface.

In addition, the 'common sense generation' criteria, which examines the ability to generate common sense, have been added to evaluate the high performance and diversity of Korean LLM models.

'Common sense generation' is a dataset built by Upstage in collaboration with Professor Lim Hee-seok's research team at Korea University, and consists of questionnaires that include a wide range of types of historical distortion, hallucination errors, morpheme errors, irregular utilization errors, and hate speech. Through this, it measures whether the results created by AI for the given conditions can conform to the general common sense that Korean speakers can have.

In other words, it is evaluated that it is possible to prevent cases such as the 'Sejong the Great's

MacBook Throwing Incident', which is evaluated as the most representative example of hallucination in Korea, through the corresponding common sense generation criteria, and it is possible to evaluate more suitable models for Korean language and history.

The Open Ko-LLM Leaderboard is expected to increase the level and competitiveness of Korean LLM research, improve the quantity and quality of Korean data, and enhance the international recognition of Korean LLM. Based on the open LLM leaderboard, it is possible to share the results of various researchers and promote joint research and cooperation, and it is expected to contribute to the improvement of performance levels and the expansion of industrial fields.

In addition, the opening of the leaderboard based on Korean data is expected to secure transparency and reliability of public research results, as well as to enhance international recognition and become a starting point for activating research on various languages.

The recent collaboration between Upstage and KT also played a role in the opening of the Open Ko-LLM Leaderboard. The two companies will join forces to expand the AI ecosystem, and the leaderboard will be operated stably through infrastructure support from KT Cloud.

Upstage's LLM model SOLA ranked first in the world in August with results that exceeded ChatGPT's benchmark score in the open LLM leaderboard operated by Huggingface. Recently, it surprised the global market again by registering SOLA as the main model of Poe, which was the standard for high-performance models, with only four LLMs from OpenAI ChatGPT, Google PaI, Meta Llama, and Anthropic Claude listed.

Upstage plans to develop high-quality LLMs that can capture Korean culture and emotions based on high-quality Korean data based on the 1T club, and contribute to the creation of an ecosystem for Korean

LLM independence.

Upstage CEO Kim Sung-hoon said, "We are very pleased to open the Open Ko-LLM Leaderboard, which will promote the competitiveness of Korean LLM and further enhance the level of research with NIA." He added, "We will continue to make efforts to expand the Korean AI ecosystem and promote its development by sharing high-quality Korean data, including the 1T club, and promoting collaboration and rapid technology dissemination through the leaderboard."