

# alt announces LHTM-OPT, a lightweight, high accuracy large language model

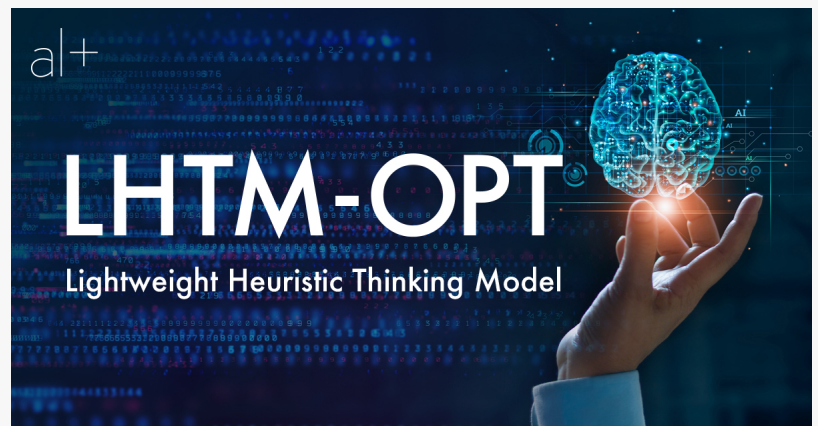
*The model achieved the highest score in the Rakuda benchmark for Japanese LLMs, making it the most accurate commercial private LLM in Japan*

TOKYO, JAPAN, October 23, 2023 /EINPresswire.com/ -- [alt Inc.](#), a Japan-based developer of AI Clones and Personal Artificial Intelligence (P.A.I.®) (head office: Minato-ku, Tokyo; CEO: Kazutaka Yonekura), is pleased to announce the development of LHTM-OPT, a new large language model (LLM) which is optimized for small GPU machines. The model has recorded the highest performance among the Japanese commercial private LLMs on the Rakuda benchmark.

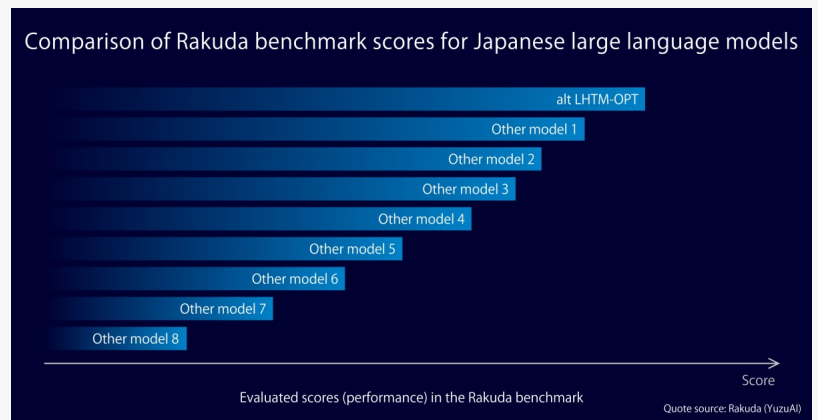
\*The highest and best performance in Japan:

Achieved top performance among domestic Japanese large language models, as evaluated by the Rakuda LLM benchmark. (As of October 19, 2023. In-house research)

Large language models with many parameters require enormous amounts of computational resources and energy, and more and more people are becoming concerned about the burden they place on the environment. For this reason, research on limited-parameter models, which have potential to achieve similar accuracy with lower energy consumption, is gaining momentum. This approach is described as "knowledge distillation" because it extracts important knowledge from LLMs to develop new useful models that use



alt Inc. announces LHTM-OPT, a lightweight, high accuracy large language model ~The model achieved the highest score in the Rakuda benchmark for Japanese LLMs, making it the most accurate commercial private LLM in Japan~



alt LHTM-OPT; Comparison of Rakuda benchmark scores for Japanese large language models

smaller computational resources. alt Inc. began actively and quickly working on the development of generative AI to solve these issues, and has successfully developed LHTM-OPT, an optimized LLM for small GPU machines.

Our classic large language model [LHTM-2](#) is set up with a huge number of parameters and a very high degree of freedom. In contrast, LHTM-OPT is a lightweight model whose number of parameters is just 5% of LHTM-2 and is optimized for practical on-premise usage on small GPU machines.

In developing LHTM-OPT, we have not only trained on open data, but also created our own sets of instruction and dialogue data to improve the accuracy. On YuzuAI Group's Rakuda benchmark (<https://yuzuai.jp/benchmark>), which evaluates Japanese LLMs, LHTM-OPT scored 1152, the highest score among the Japanese LLMs as of October 19, 2023. LHTM-OPT also achieved Japan's highest score for a commercial model on the 8-task average of the JGLUE benchmark (Japanese General Language Understanding Evaluation). (LHTM-OPT's average score for the four major JGLUE tasks was 75.08.)

The achievement of these record high scores with an under 10B parameter LLM demonstrates our skills and knowledge on LLM training, data creation and model optimization. We will continue to improve the performance and versatility of LHTM-OPT by continuously evaluating the model against various benchmarks and fine-tuning its performance. We'll also use the model as a basis for implementing a number of applications to handle problems requiring advanced LLM tuning techniques, such as automatic question answering, automatic summarization of minutes, information extraction, conversation understanding, and data organization and creation for predictive analysis.

□The significance of the lightweight LLM achieving high accuracy and the possibility of providing value to clients□

□ Significance: Highest level of "knowledge distillation" technology in Japan, achieving high precision even with a low number of parameters

□In the learning and evaluation process, including this "knowledge distillation," we practice consistent planning, technical development, and operation, from data planning to creation, evaluation and securing of GPU resources, and optimization of the learning process. As a result, we achieved a lightweight and highly accurate model, showcasing the quality of alt's technology in building the foundation for more practical generative AI products.

□ Value provided: Ease of customization

□Our clients can fine-tune the newly developed lightweight and highly accurate model to their specific fields. By customizing LHTM-OPT based on information specific to various fields such as finance and pharmaceuticals, manufacturing and real estate, our client can provide expertise and answer accuracy equivalent to LLMs with overwhelmingly large numbers of parameters.

□ Value provided: GPU cost savings

Our clients' implementations will also be optimized for running costs, which is a challenge when building services based on large language models. By operating lightweight and highly accurate models, we can help them maximize the cost-effectiveness of continued use.

We aim to become the "OpenAI of Asia," by championing world-class technology, providing the highest-quality solutions to our customers, and promoting initiatives that contribute to improving the labor productivity of Japanese companies.

For more information on projects utilizing LLMs, please reach out to the alliance contact point below.

For inquiries about LHTM-2/LHTM-OPT/GPT and other large language models solutions  
<https://alt.ai/aiprojects/gpt/>

About alt Inc.

Established in November 2014, alt Inc. is a venture firm with the mission of freeing humankind from non-creative/unproductive labor through the creation of P.A.I.® (Personal Artificial Intelligence) and AI clones. We also develop and provide SaaS products such as [AI GIJIROKU](#), which utilizes voice recognition technology born from the development of AI dialogue engines. As of September 2023, our cumulative funding has reached over 8 billion yen.  
<https://alt.ai/en/>

<Media Inquiries to:>

Misako Nishizawa (Media Relations)  
e-mail [press@alt.ai](mailto:press@alt.ai)

<Alliance Inquiries to:>

We provide AI solutions and support regardless of genre, including IT, finance, construction, logistics, media, manufacturing, retail, and service industries.  
Please feel free to contact us.

Katsuya Asai (AI Solutions Business Department)  
e-mail [gptsolutions@alt.ai](mailto:gptsolutions@alt.ai)

Misako Nishizawa  
alt Inc.  
+81 3-6455-4677  
[press@alt.ai](mailto:press@alt.ai)

Visit us on social media:

[Facebook](#)

[Twitter](#)

[LinkedIn](#)

[YouTube](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/663570384>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2023 Newsmatics Inc. All Right Reserved.