"exaBase Generative AI" Supports Multiple Large Language Models and Patent Technology

to Optimize the Use of Switching between LLMs

～Plans to Support "Claude" by Google, "Gemini" by Google, "tsuzumi" by NTT,

Expanding the Application Fields and Uses of Generative AI by Utilizing Multiple LLMs～

The Company group is announced that at its business meeting held today, The Company's subsidiary, Exa Enterprise AI, Inc. has decided to support multiple Large Language Models (LLM) in its corporate ChatGPT service, exaBase Generative AI.

In addition to the GPT series of OpenAI, the system will support Claude3 by Anthropic, Gemini by Google, and tsuzumi by NTT Group starting in May 2024. In addition, exaWizards has developed migration technology to absorb differences in different LLMs, which will be incorporated into exaBase Generative AI. This technology has been patented.

Through these efforts, the ExaWizards Group will realize a more competitive usage environment in terms of applications and situations, and expand the fields of application of generative AI.

１．Expansion of "exaBase Generative AI" service

In response to the rapid spread of generative AI, a number of global technology companies have begun to develop and offer LLMs. Each LLM developed by various companies has its own strengths, and by selectively switching between them depending on the application, it can be used to improve productivity in the target business or application.

■Selecting LLMs and utilizing highly accurate and different opinions

Specific applications of multiple LLMs are expected to include the following cases.

1. Expanding the range of idea generation, finding appropriate expressions, and reducing omissions in key points

2. Output of analysis reports, etc., by combining GPT's ability to analyze and plan things and Claude's strength in writing

3. Enhancement of Japanese language support by using tsuzumi, a domestically produced LLM

4. Cross-checking of output information to determine the presence or absence of hallucination and bias

■Migration is achieved through the use of improved prompts.

The output generated in response to a prompt can vary significantly depending on the LLM used. This variability becomes particularly problematic when the prompt is embedded in a system or when it is used

frequently by many users as a standard template. Additionally, there is a risk that the primary LLMs being used could become unavailable for various reasons.

To address these issues, ExaWizards has developed a technology that enables the migration of LLMs. This technology features two key components: (1) an evaluation of the differences between two LLMs for a given prompt, and (2) the generation of "improved prompts" that ensure compatibility with the new LLMs. Our assessments have confirmed that this migration process is effective for a wide range of prompts.

■ Regarding LLM Migration Technology:
Initially, we will offer our LLM migration technology in conjunction with consulting services by ExaWizards. This will be in response to requests from clients currently utilizing our AI generation services. We also aim to integrate this technology's functionality directly into exaBase Generative AI.

■ Patent Overview for LLM Migration Technology:
This patent enables a computer to adjust prompts with the aim of minimizing variations in prompt responses when transitioning to a different LLM. With this patented technology, it's possible to migrate LLMs while ensuring consistency in outputs.

Patent number: 7471044
Registration date: April 11, 2024
Patent holder: ExaWizards, Inc.

2．Future Plans
・New support for LLMs in exaBase Generative AI
We plan to support new LLMs within exaBase Generative AI, starting with Claude3 in May 2024. Following that, we will add support for other models like Gemini and tsuzumi*.
・Regarding LLM Migration Technology:
Initially, we will offer our LLM migration technology in conjunction with consulting services by ExaWizards. This will be in response to requests from clients currently utilizing our AI generation services. We also aim to integrate this technology's functionality directly into exaBase Generative AI.

3．The amount and details of special expenditure for the commencement of the business concerned
There are no plans at this time to make any special expenditures, and we will proceed in the normal course of business operations.

４．Impact on Business Performance

The impact of this business on The Company's business performance for the fiscal year ending March 31, 2025 is under scrutiny, but we expect that this business will contribute to improving The Company's business performance in the mid-to-long term.

The Company plans to announce its fiscal year financial results on May 14, 2024, at which time it will disclose its earnings forecast in light of the impact of this business on its performance.

-End-