Using Large Language Models to Achieve DX

Developing Chat-IHI, an environment based on ChatGPT, and using it in business

The IHI Group is making use of generative AI, including large language models (LLMs), to achieve digital transformation (DX). This article introduces the three phases of the IHI Group's LLM utilization and the specific initiatives involved in each of these phases.

Image of using Chat-IHI

Introduction

The IHI Group is promoting digital transformation (DX) initiatives to achieve business process reform with the aim of improving operational efficiency and shortening lead times as well as business model reform with the aim of solving social issues. In terms of business process reform, the IHI Group aims to achieve labor savings, particularly by making use of data and digital technology. Generative AI, which includes the large language models (LLMs) and has attracted attention in recent years, is an important technology for achieving labor savings because it can perform creative tasks such as document preparation and planning that have traditionally been performed by humans.

ChatGPT is an interactive AI service that generates natural sentences similar to human conversation based on LLM. The IHI Group is advancing its use of LLMs through the following three phases. In phase one, we used ChatGPT, a publicly available LLM that has been trained only on open-source information, as-is in our business. In phase two, we have developed and are using a system that refers to internal business data. In phase three, we will build a system to run an LLM for each business area. In this article, we will introduce the IHI Group's initiatives for utilizing LLMs.

Developing a ChatGPT environment

ChatGPT is an artificial intelligence chatbot that OpenAI released in November 2022. It is a conversational generative AI service based on a generative pre-trained transformer (GPT), a type of LLM, that is capable of conducting natural-sounding conversations as if speaking to a human. This text generation capability makes it possible to generate sentences for things like emails and notifications, to organize survey results, to translate and summarize sentences, to create minutes from meeting notes, and to write programming code. As a result, there are rising expectations regarding its ability to support and reduce the time spent on work that was previously performed by humans. The IHI Group also began preparations to introduce ChatGPT internally within the Group in April 2023 to improve operations.

Since ChatGPT is a service that is publicly available via the Internet, if it is used without modifications, there is a risk of information leakage, such as the possibility of the LLM being trained on internal information uploaded by users and then providing that information to external users. To minimize this risk, we developed Chat-IHI, an internal ChatGPT service that uses Microsoft's Azure OpenAI Service on the IHI Group's private cloud. Though Chat-IHI is an LLM service for the IHI Group, in phase one, we did not customize it for IHI; rather, we used the model provided by Azure OpenAI Service as is.

Azure OpenAI Service is a generative AI service for business use based on OpenAI's GPT, delivered via the Microsoft Azure cloud. In general, this service allows Microsoft to record and utilize user usage logs, but we developed Chat-IHI after taking the steps necessary to prevent such logging.

At the end of June 2023, Chat-IHI was made available to approximately 20,000 employees of the IHI Group companies in Japan. Overseas, it will be gradually rolled out in countries and companies where the necessary infrastructure is in place. At the same time, we launched the Chat-IHI portal site to consolidate and disseminate information on using Chat-IHI in business, including notes on use, instruction manuals, and FAQs. While Chat-IHI is useful for improving operational efficiency and generating new ideas, there is a risk that using it may violate laws or infringe the rights of others, depending on the content entered and on how the generated output is used. Therefore, we have published guidelines to inform employees about what kinds of data can be entered into Chat-IHI and how to use the output results appropriately.

These guidelines are based on the "Guideline for the Use of Generative AI" published by the Japan Deep Learning Association. However, as the IHI Group is involved in the nuclear power and national defense businesses, we needed to give particular consideration to how to handle technical information related to national security. Therefore, in addition to consulting with internal departments involved in legal, economic security, and other matters, we also consulted with external experts, such as lawyers, to develop the guidelines. These guidelines include points that users should keep in mind when entering data, such as "Never enter personal information or confidential information from your own department into Chat-IHI," and points to keep in mind when using generated output, such as "Since the output generated by Chat-IHI may not be factually accurate, do not blindly trust the generated output, and be sure to obtain evidence and supporting information."

Chat-IHI screen

Chat-IHI portal site screen

Using Chat-IHI to improve operations

After releasing Chat-IHI, 3,600 people used it in the first week, indicating a high level of interest. Here, we will introduce the first phase in the IHI Group's use of LLMs: our initiative to use a publicly available LLM in our business.

This section explains the operational improvement campaign, which was an activity in the first phase in using LLMs. After its introduction, Chat-IHI was used to improve individual work, but improvement activities at the workplace level were not fully realized. To encourage the habitual use of Chat-IHI in each workplace, we conducted a campaign to raise awareness about using Chat-IHI. This campaign was conducted from August to September 2023 and targeted about 200 employees in an information system division.

The campaign involved using Chat-IHI to improve daily operations and then reporting the results within the division. We asked that the reports should include not only examples of use but descriptions of the effects of the improvements made, such as reductions in the number of working hours. Over the course of three weeks, several dozen cases were reported, including document preparation assistance, coding assistance, idea generation, and analysis and proposal assistance.

In one case in which the customer service group used Chat-IHI to analyze a written questionnaire, the time required for analysis, which would have taken several hours if done by a person, was reduced to a few minutes. By entering the content of the questionnaire and the instruction "Please perform a trend analysis" into Chat-IHI, the trends were displayed in bullet points, thus greatly reducing the burden of understanding the questionnaire trends.

In other cases as well, we could expect a reduction in working hours of around 50% to 90%. Use of Chat-IHI enabled us to confirm its effectiveness in reducing the number of working hours and assisting in idea generation. In addition, use of Chat-IHI in the workplaces targeted by the campaign allowed us to promote the use of Chat-IHI in business.

Moreover, we collected useful examples of instructions (prompts) given to Chat-IHI by users inside and outside the Group, and we posted specific methods for writing prompts for each category of use on the portal site, thus introducing useful methods for making operational improvements.

Our initiative to leverage internal data

This section explains the second phase of the IHI Group's use of LLMs, namely to link LLMs to business data.

To use Chat-IHI to achieve internal business process reform, one possibility is to ensure Chat-IHI's responses reflect internal data (e.g., internal regulations). In addition, we are considering using Chat-IHI to identify potential risks in safety risk assessments conducted prior to the start of work at construction sites by referencing past cases of internal and external disasters. To achieve this, we have been working to develop a system that responds with business knowledge using a technology called retrieval augmented generation (RAG), which links external databases and information sources.

Also, because GPT and other LLMs sometimes generate hallucinations (the phenomena of generating information not based on facts), it is necessary for users to verify the accuracy of the response. As a countermeasure against hallucinations, this system provides links to related documents alongside the responses, making it easier for users to verify the response accuracy.

We were able to confirm that using RAG makes it possible to link generative AI responses to internal data while reducing the impact of hallucinations, which are a weakness of LLMs.
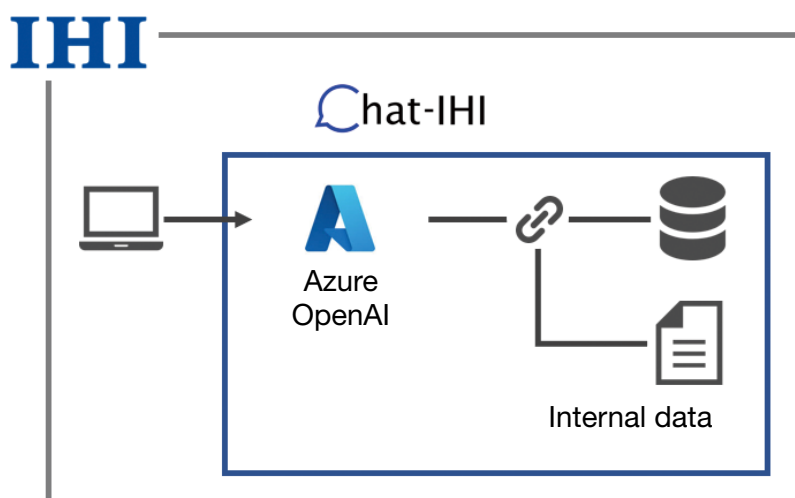
Configuration diagram for a system using RAG

Users can review the original text of the internal knowledge used to generate the answer.
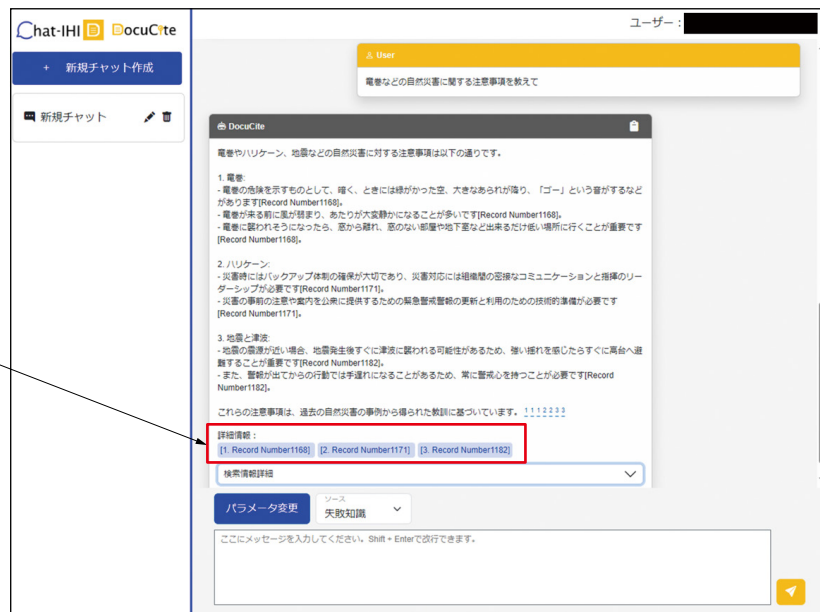
Business knowledge answering system using RAG

Developing a system for operating LLMs for each business area

This section explains the third phase of the IHI Group's use of LLMs: an initiative to develop a system for operating LLMs for each business area. Since the IHI Group is a group of companies operating in four business areas, it is necessary to manage confidential information, such as by ensuring that confidential information in one business area cannot be viewed by another business area, while sharing useful data across business areas. One solution to this problem is to prepare an environment for operating LLMs in each business area, but since preparing a cloud environment for each business area is expensive, we are considering the use of lightweight LLMs that can be operated with a small amount of computing resources. Because lightweight LLMs have fewer parameters, they do not require large servers like those in cloud environments.

Also, because lightweight LLMs have a small number of parameters, their accuracy is inferior to models with a large number of parameters, but they are easier to fine-tune for business data and business tasks than normal LLMs, so they can be expected to accelerate application of LLMs into business. In the future, we will continue to work to create a lightweight LLM environment and apply it in our businesses.

Conclusion

To use LLMs for achieving DX, the IHI Group is taking a three-phase approach: using publicly available LLMs in business, linking LLMs with internal data, and building a system to operate LLMs for each business area.

In the first phase, we conducted an operational improvement campaign and shared examples of how LLMs are being used across the Group to promote the use of LLMs in business. In the second phase, we have been using RAG to develop a system that enables us to search internal data while taking measures against hallucinations.

In the future, we will work on the third phase, namely developing a system to use LLMs for each business area, as well as actively engage in using LLMs to address social issues, not just to reform business processes.