

# RAG で最適化した生成 AI による HPC ユーザ向けサービスの実現

三上和徳<sup>1,a)</sup>, 中村宜文<sup>1,b)</sup>, 庄司文由<sup>1,c)</sup>

1) 理化学研究所 計算科学研究センター

a)kazunori.mikami@riken.jp, b)nakamura@riken.jp, c)shoji@riken.jp

## Realization of HPC User Services Using Generative AI Optimized with RAG

Kazunori Mikami<sup>1,a)</sup>, Yoshifumi Nakamura<sup>1,b)</sup>, Fumiyoshi Shoji<sup>1,c)</sup>

1) RIKEN Center for Computational Science

### 概要

理化学研究所計算科学研究センター（R-CCS）ではスーパーコンピュータ「富岳」のユーザから寄せられる様々な技術的質問や要望へのサポートを行う「富岳サポートサイト」におけるサービスの一環として、2024 年度から生成 AI によるサービスを加え、ユーザ自身による迅速な自己解決を実現するための取り組みを推進している。本稿では R-CCS が生成 AI をサービスに採用した経緯、RAG を応用した最適な生成 AI サービスの構築、得られた効果などに関して報告を行う。また、同じ生成 AI 技術を応用してサービスを開始した HPCI 利用報告書の閲覧支援サービスについても紹介をする。

## 1 ユーザをサポートするサービス基盤

### 1.1 サービス形態の段階的な高度化

サービス形態が段階的に高度化してきた経緯について説明

1. メールベースでのサポート
2. 富岳サポートサイトの開設  
(キーワード: Zendesk を基盤とするチケットサービスへの移行)  
(キーワード: ウェブブラウザ UI、チケットサービス)
3. 生成 AI AskDona による自動回答サービスの追加  
(キーワード: 生成 AI、RAG、自動回答)

## 2 富岳サポートサイトへの生成 AI の導入

論点: 導入背景: ユーザ側でのハードル、運用側でのハードル

論点: ユーザ自身による問題解決の促進

論点: 生成 AI 提供者数社による概念検証の実施

論点: 入札による業者決定。機能強化・改善要望に機動的に対応する業者が選定されたことは重要なポイントであった

## 3 生成 AI AskDona

論点: AskDona の構成概要

論点: RAG のメリット (回答正解率の高さ、ハルシネーションが起きにくい構成)

論点: 富岳サポートサイト専用の知識データベースの構築

論点: チケットサービスと生成 AI チャット機能の統合

論点: 個人情報の不所持方針

## 4 生成 AI 導入の効果

論点: 発行チケット数の変化

論点: ユーザからのフィードバック

論点: ユーザの利用方法の変化

論点: 課題: ユーザフィードバック率の低さ

論点: 課題: 人手で対応するチケットサービスの満足度 V.S. 生成 AI サービスの満足度

## 5 HPCI 利用報告書の閲覧支援 AI サービス

論点: 同じ AI 技術の水平展開

論点: HPCI 利用課題の検索、論点整理、調査・比較

論点: JHPCN 成果報告書の閲覧機能を追加

%参考文献

## 参考文献

- [1] 雑誌の場合：著者名、タイトル、雑誌名 巻、号、  
ページ、発行年 .
- [2] 書籍の場合：著者名、書名、参照ページ、発行所、  
発行年 .