

# RAG+生成AIによる富岳ユーザーサポート

中村宜文

理化学研究所計算科学研究センター

2025年6月27日

PCクラスターワークショップ in 柏 2025



理化学研究所  
庄司文由、三上和徳、中村

GFLOPS  
盛本マリア、鈴木 亮祐

# 「富岳」ユーザー支援における課題と生成AI導入の狙い

- 背景

- 技術文書やFAQが多様なフォーマット、その配置場所が分散
- 複雑な問い合わせに対して、利用者が**必要情報へ迅速にたどり着けない**
- 従来の検索機能では**意味的な理解や横断的な検索が困難**
- 問い合わせ対応の人的負荷が大

- 目的

- 利用者自身による**問題の自己解決率の向上**
- サポート業務の効率化とスケーラビリティの確保
- **自然言語による即時質問対応の実現を通じた利便性向上**

- 導入

- 生成AI+RAGチャット型アシスタントの有効性を検証
  - 約半年間、複数社製品について、評価、比較、PoC
  - GFLOPS「**AskDona**」を導入決定

<https://www.gflops-ai.com/>

<https://www.askdona.com/>

- **2024年7月9日（利用者向け公開、プレスリリース）**

導入プレス記事



# 富岳サポートサイトにおける「AI」による支援



富岳のマニュアルやナレッジベース等を学習した生成AIアシスタント **AskDona** が富岳の利用に関する質問に迅速かつ的確にチャット形式で回答

高度AI検索

## AskDonaのデータベース (随時拡充)

- 富岳利用手引書 (システム利用、プログラミングガイド)
- マニュアル (言語、FEFS、ジョブ運用ソフトウェア)
- A64FXチューニング事例
- 富岳サポートサイト (zendesk) のFAQ記事
- その他：基本仕様やシステム構成など、運用情報などのお知らせ、理研が提供するソフトウェアのユーザガイド、講習会資料

# AskDonaのユースケース

生成AIアシスタント - AskDona

会話型ジョブで複数のノードを利用したいときはどのように設定したらいいですか？

会話型ジョブで複数ノードを利用するには、以下の手順を参考にしてください。

1. **pjsubコマンドの使用:** 会話型ジョブを実行するには、`pjsub` コマンドに `--interact` オプションを指定します。このオプションを使用することで、対話的にジョブを実行することができます。
2. **ノード数の指定:** `-L "node=X"` オプションを使用して、必要なノード数を指定します。ここで `X` は使用したいノードの数です。
3. **MPIプロセスの指定:** `--mpi "proc=Y"` オプションを使用して、必要なMPIプロセス数を指定します。ここで `Y` は使用したいプロセスの数です。
4. **ジョブの実行時間の指定:** `-L "elapse=HH:MM:SS"` オプションで、ジョブの最大実行時間を指定します。

以下は、複数ノードを使用する会話型ジョブの実行例です：

```
pjsub --interact -g グループ名 -L "node=2" -L "rscgrp=int" -L "elapse=1:00:00" --mp
```

このコマンドは、2ノードを使用し、32のMPIプロセスを実行する会話型ジョブを1時間の実行時間で行います。

詳細な手順やオプションについては、[利用手引書](#)をご参照ください。

Donaの回答が役に立ちましたか？フィードバックをいただくと嬉しいです。

コピー 良い回答 悪い回答

## 回答改善のためのフィードバック機能

MPIプログラムで、メモリ使用量を削減する方法を教えてください

MPIプログラムでメモリ使用量を削減するためには、以下の方法を考慮することができます。

### 1. MCAパラメータのチューニング:

- MPIライブラリが使用可能なメモリ使用量を制限するために、MCAパラメータを調整します。具体的には、`common\_tofu\_memory\_limit` を1以上の整数値に設定することで、メモリ使用量の制限を有効にできます。この設定により、MPIライブラリは内部的に各MCAパラメータを自動的にチューニングし、指定されたメモリ使用量の範囲で動作するようになります。
- 詳細は[こちらの資料](#)をご参照ください。

### 2. 通信モードの選択:

- 高速型通信モードと省メモリ型通信モードを使い分けることができます。省メモリ型通信モードでは、メモリ使用量を抑えつつ通信を行うことが可能です。このモードには、Medium受信バッファとShared受信バッファを使用する方式があります。
- 詳細は[こちらのマニュアル](#)をご参照ください。

### 3. 動的コネクションの利用:

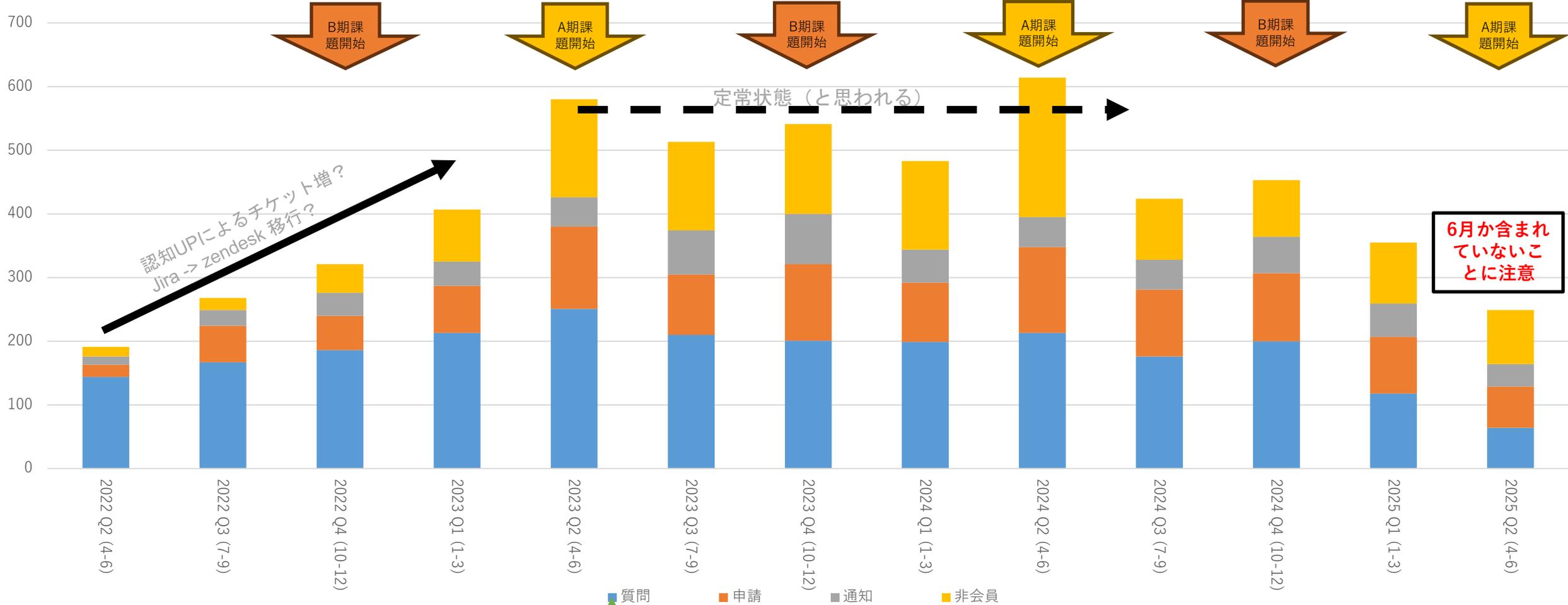
- 通信相手となる各プロセスと初めて通信を行う時点でメモリを確保する動的コネクション方式を採用することで、メモリ使用量を抑えることができます。

これらの方法を組み合わせることで、MPIプログラムのメモリ使用量を効果的に削減することが可能です。ぜひ試してみてください。

Donaの回答が役に立ちましたか？フィードバックをいただくと嬉しいです。

# 富岳サポートサイトのチケット数の推移

新規チケット発行（四半期）



6月が含まれていないことに注意

新チケットシステム運用開始  
2022/05/11

大規模FAQ整備  
2023/03

質問  
↑  
AIサポートにより減少を見込むチケット

サポートサイトデザイン改善  
2024/03/22

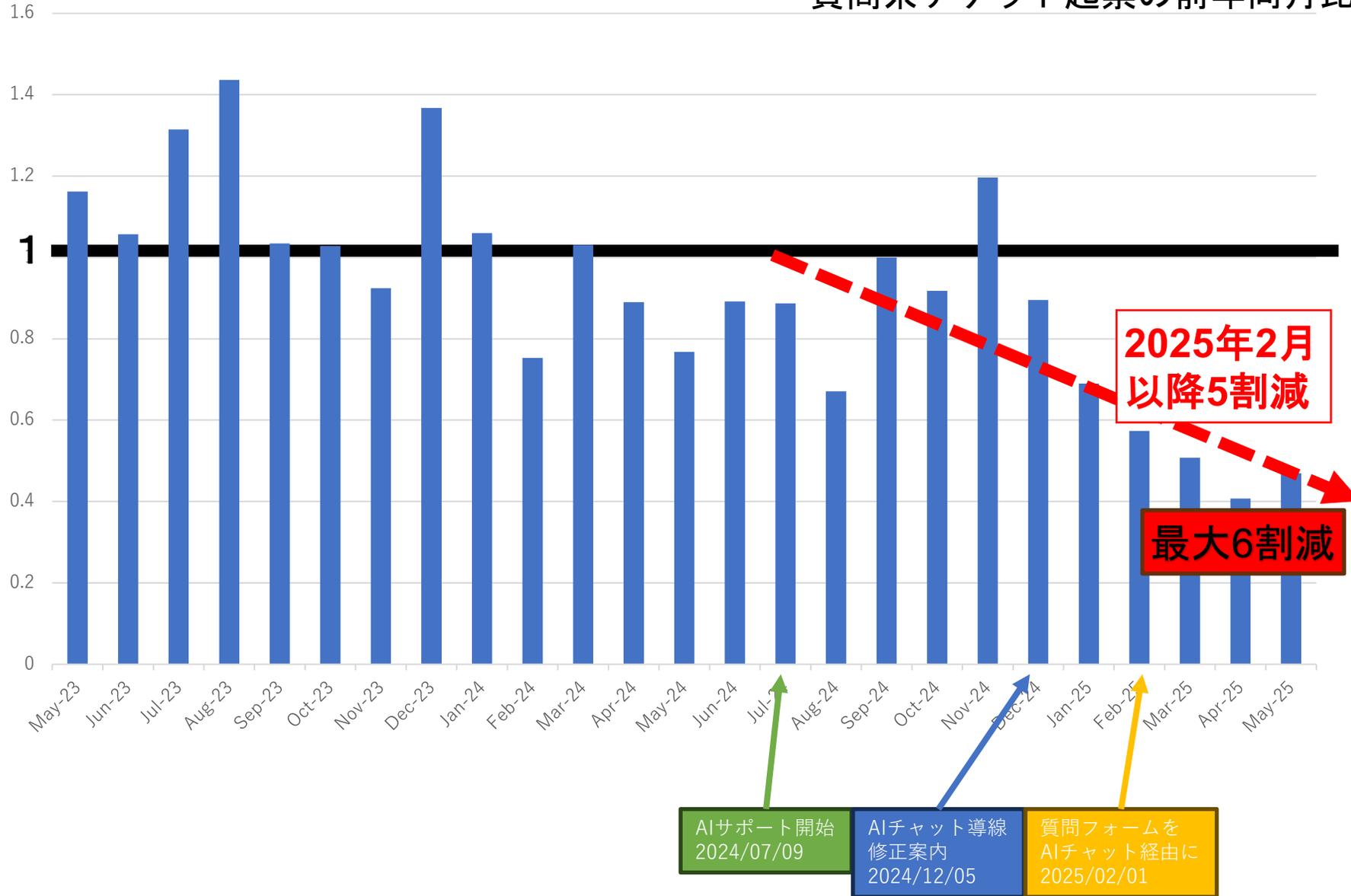
AIサポート開始  
2024/07/09

AIチャット導線修正案内  
2024/12/05

質問フォームをAIチャット経由に  
2025/02/01

# RAG+生成AIによる効果

質問系チケット起票の前年同月比



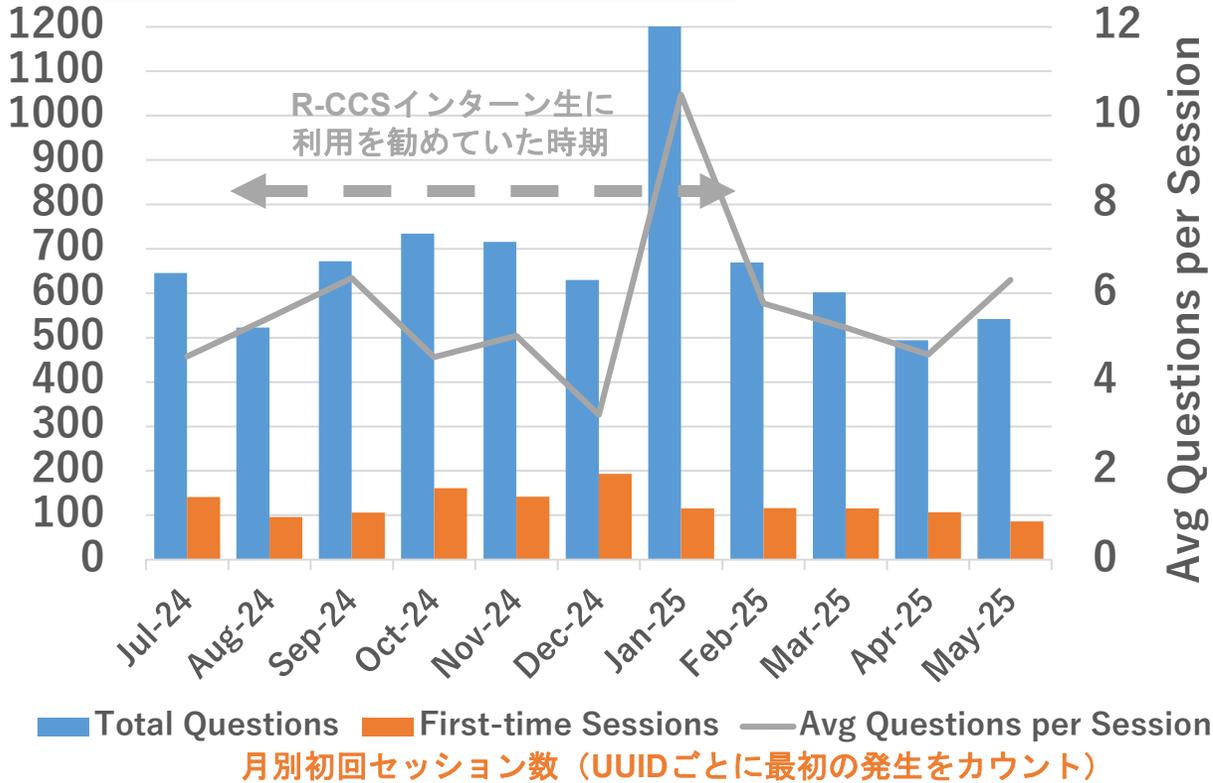
富岳ユーザ数は増加傾向  
(Ref. 『富岳年報』)

年度	アクティブユーザ数	利用者数
2020	約170/日	1798
2021	約330/日	2321
2022	約355/日	2916
2023	約395/日	3521

※2021/03/09共用開始

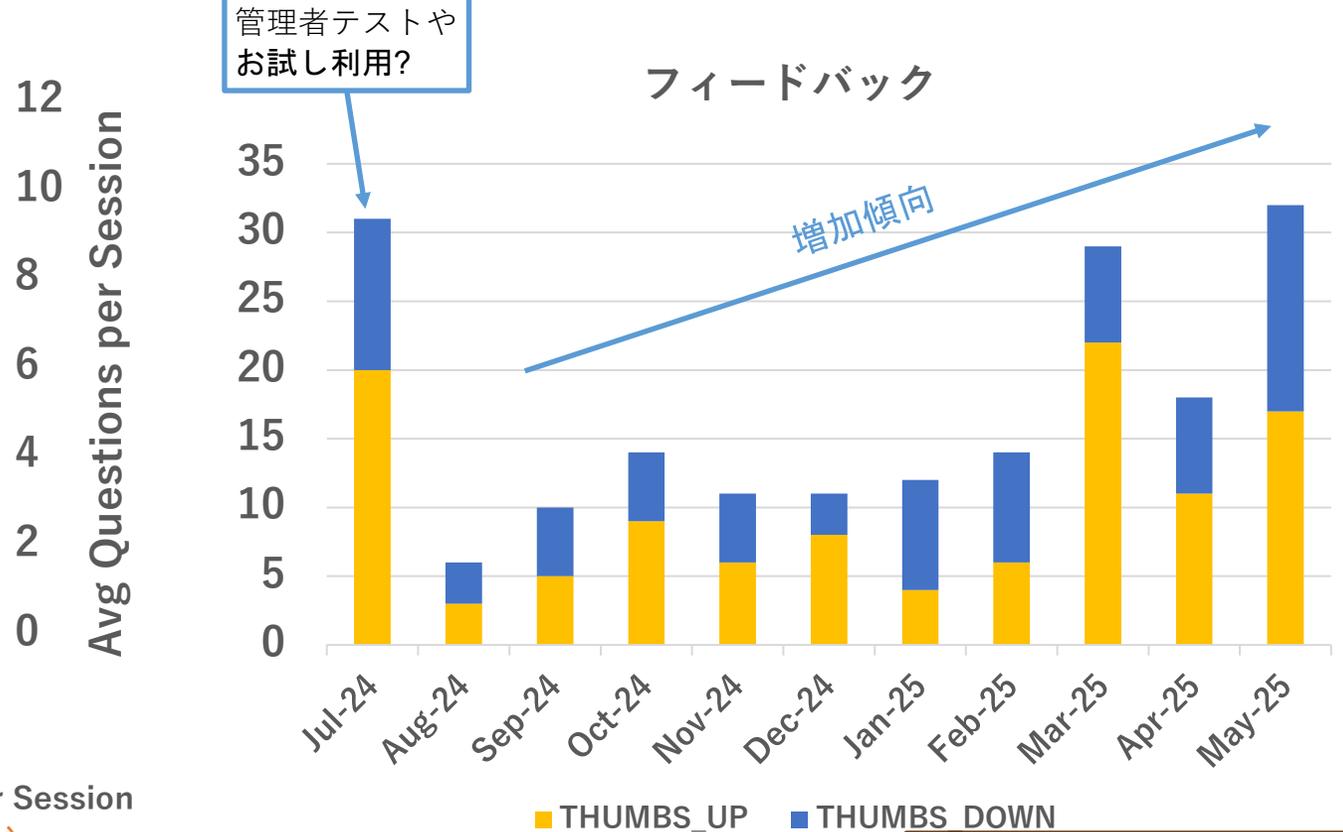
# AskDona利用実績

質問数/月: 676 (577)  
 セッション数/月: 125 (106)  
 セッション当たり質問数/月: **5.7 (4.3)**  
 ※()内は2025年2月以降の平均



- 継続利用と対話利用の定着
- 活用の効率化？

セッション当たりフィードバック率は増加傾向  
 THUMBS\_UP : THUMBS\_DOWN = 10:7

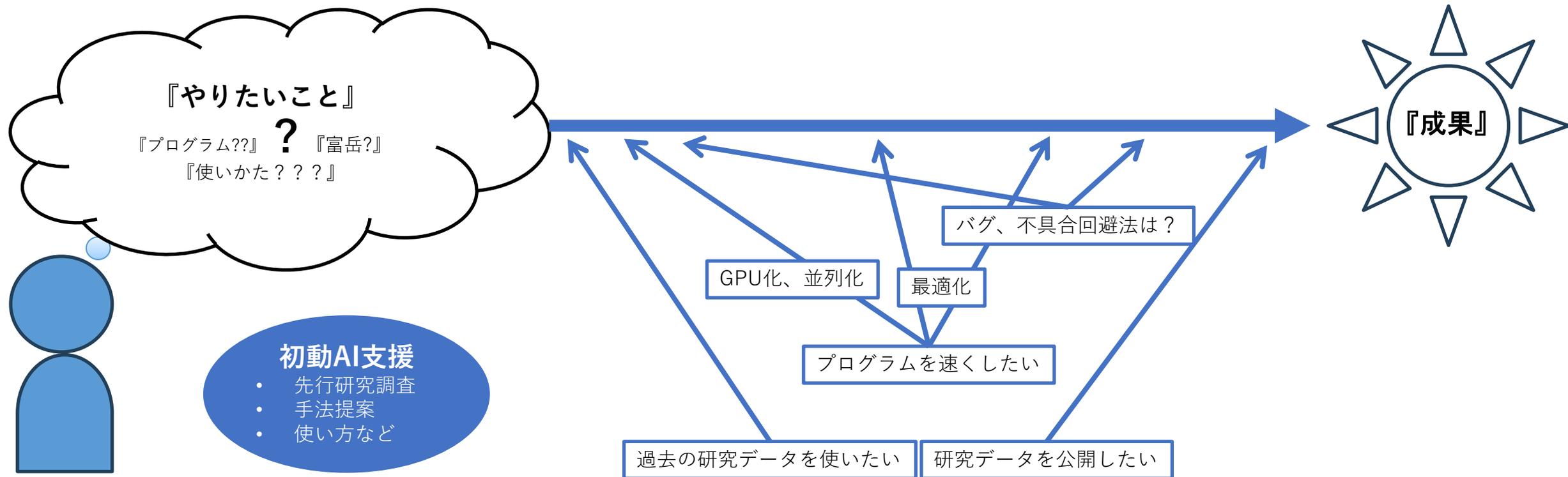


- フィードバックの増加
  - 自然な対話による動機付け、期待や要求の高まり？

最近はより複雑なクエリが増えている  
 (個別のチャット履歴から)

# 今後目指したいところ

AIを「対話」から「解決」のパートナーへ進化させ、HPCにおけるユーザサポートの質とスピードを革新



『初めての一步』から『成果』まで、あらゆるシーンでユーザを即時支援する『AI』

# まとめ

- RAG+生成AIによる富岳ユーザサポートについてご紹介しました
  - 導入まで
  - AskDona
  - 富岳サポートサイトでの「AI」支援
    - 実例
    - チケット推移・効果
    - 利用実績
- 今後目指したいところ
- 直近の課題
  - AskDona RAG バージョンアップによるさらなるサポート対応効率化
    - RAGデータベース拡大、拡充など（生成AIも活用した自動化）

**ご清聴ありがとうございました**

