



# 後期PBL成果発表

## 鉄道ナレッジシステムの高度化

CS25手島班 二宮・中川



# 背景と目的

前期のRAGを活用した鉄道ナレッジシステムでは、精度に課題が残った

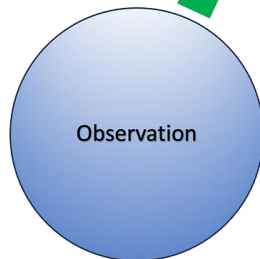
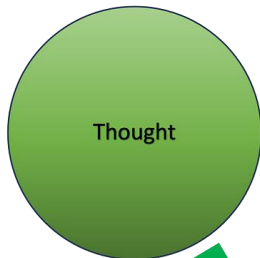
前期PBLの成果を拡張し、鉄道技術基準や鉄道事故の分析・調査・改訂を支援することで**業務負担軽減を目的とする、鉄道ナレッジシステム**を開発した。



# 着眼点と基本方針

## ReActによる鉄道ナレッジシステムの精度向上、作問と採点の自動化

Question



If Finish Action



Finish



### ReActによる精度向上

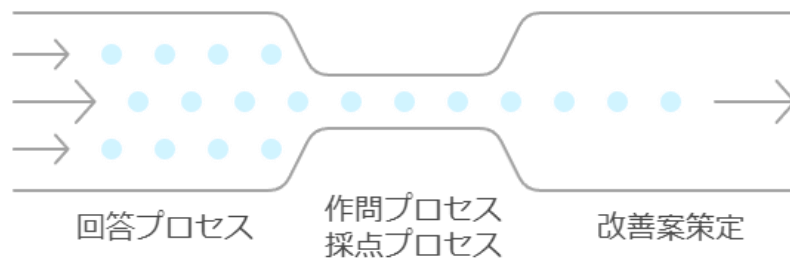
言語モデルが質問から思考（Thought）を生成、行動（Action）と観察（Observation）のループ処理により改善。

言語モデルが完了と判断すると、最終的な回答を出力。

### PDCAサイクルの高速化

前期開発では手作業で作問と採点  
その結果、PDCAの速度が低下。

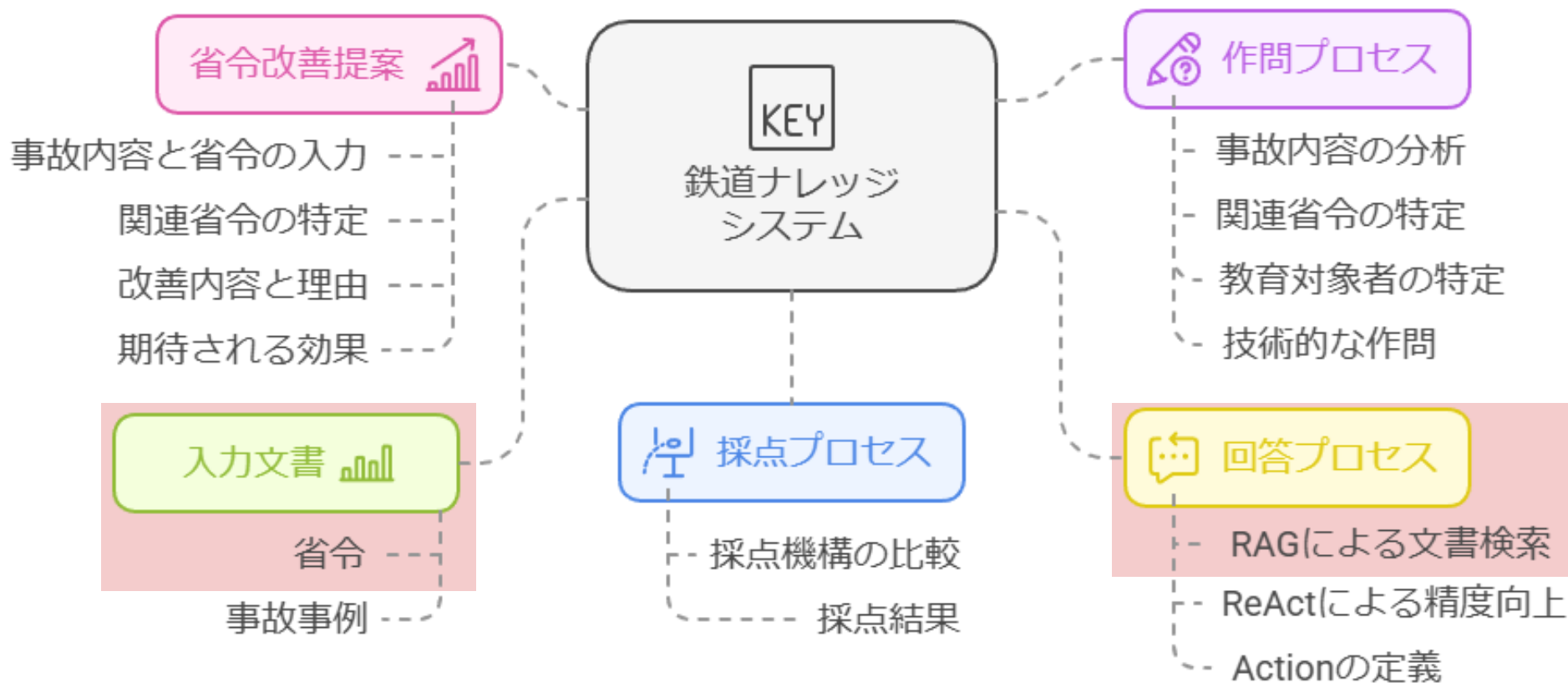
そこで言語モデル活用による自動化



# 開発内容 | 全体像

## 鉄道ナレッジシステムの全体像（前期での開発個所を赤で強調）

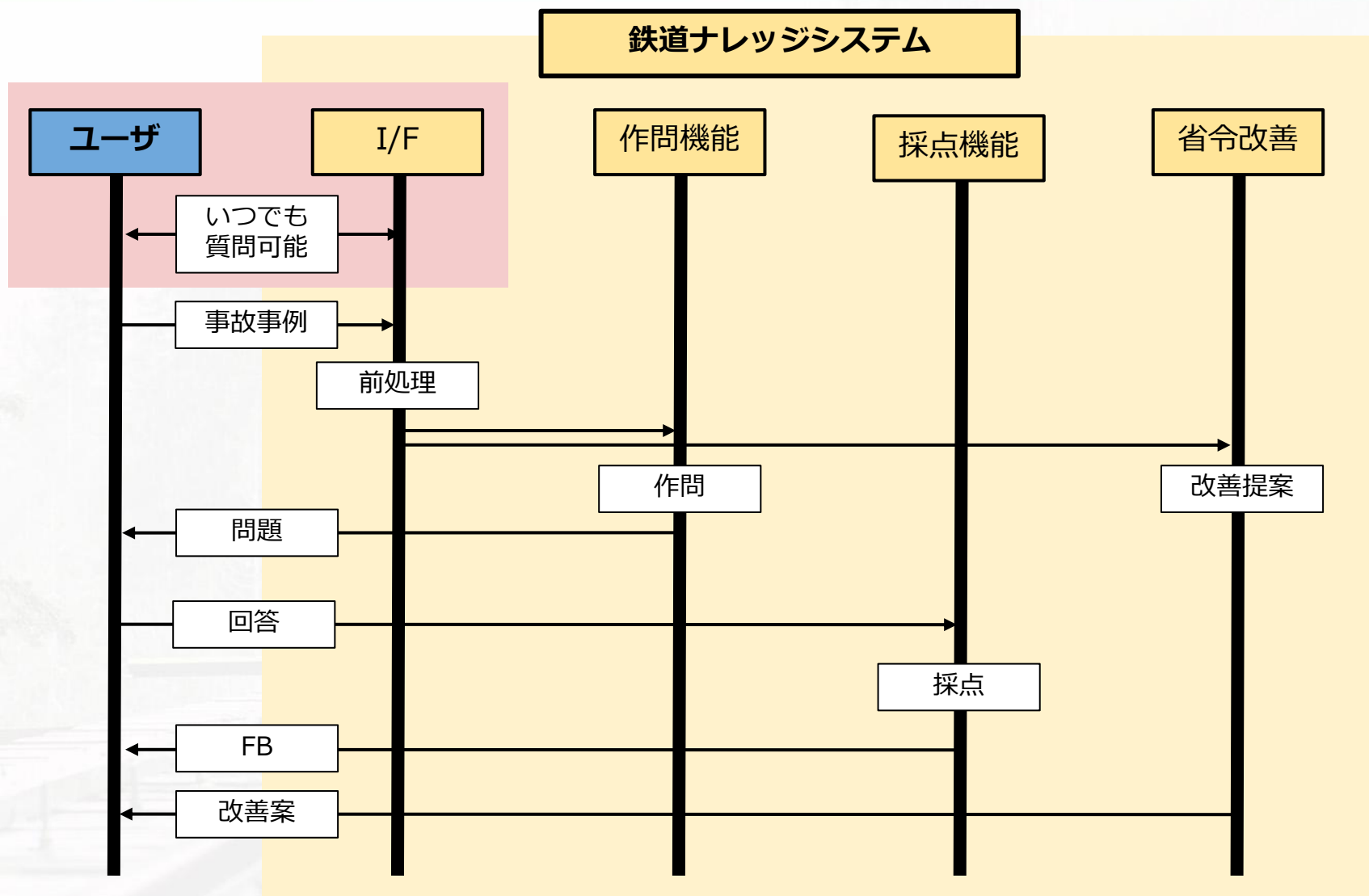
前期から、ReActの導入による回答精度向上や、作問・採点の自動化、事故報告から省令の改善提案機能を加え、**鉄道ナレッジシステムを拡充**

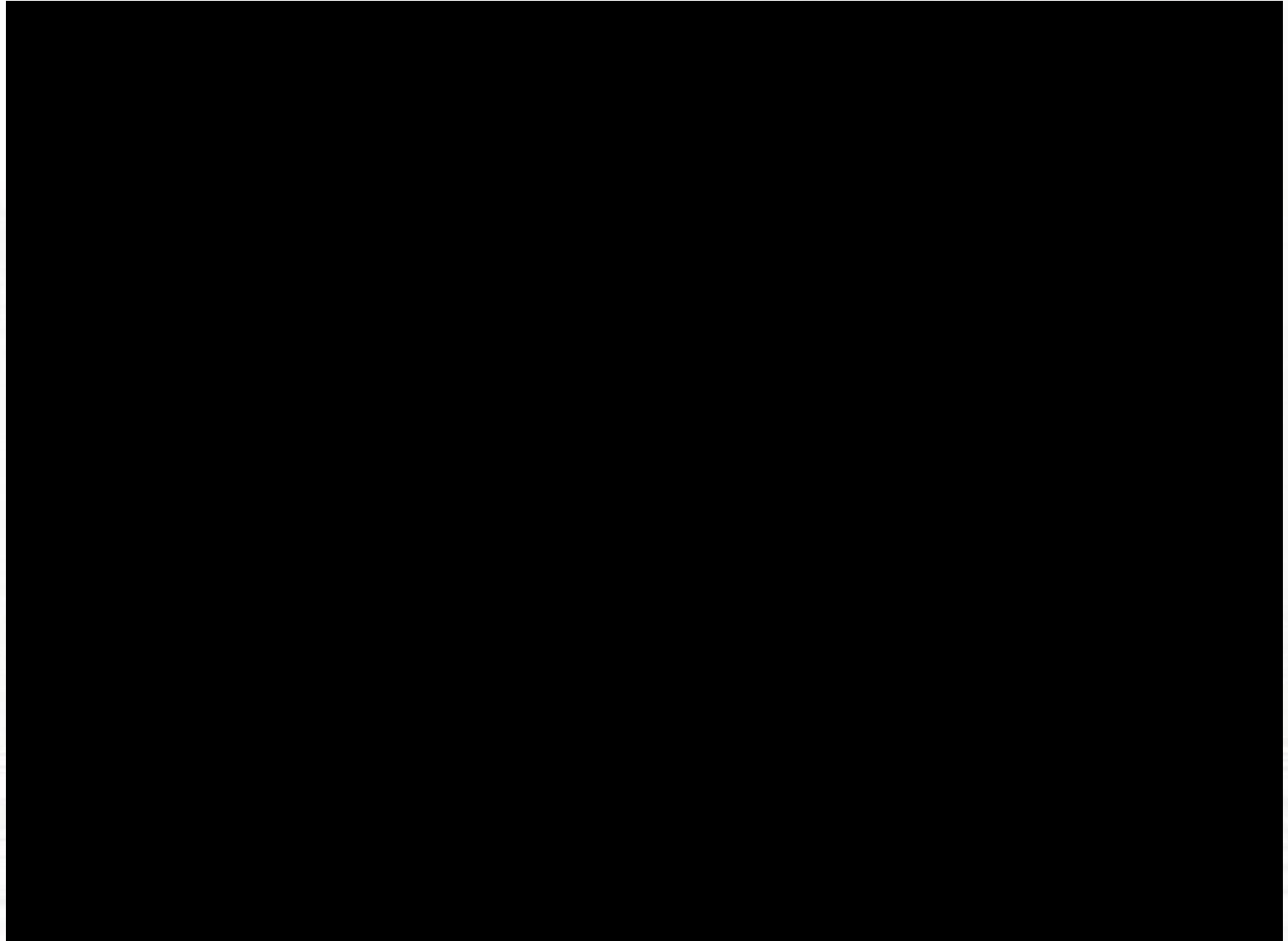




# 開発内容 | システム構成図

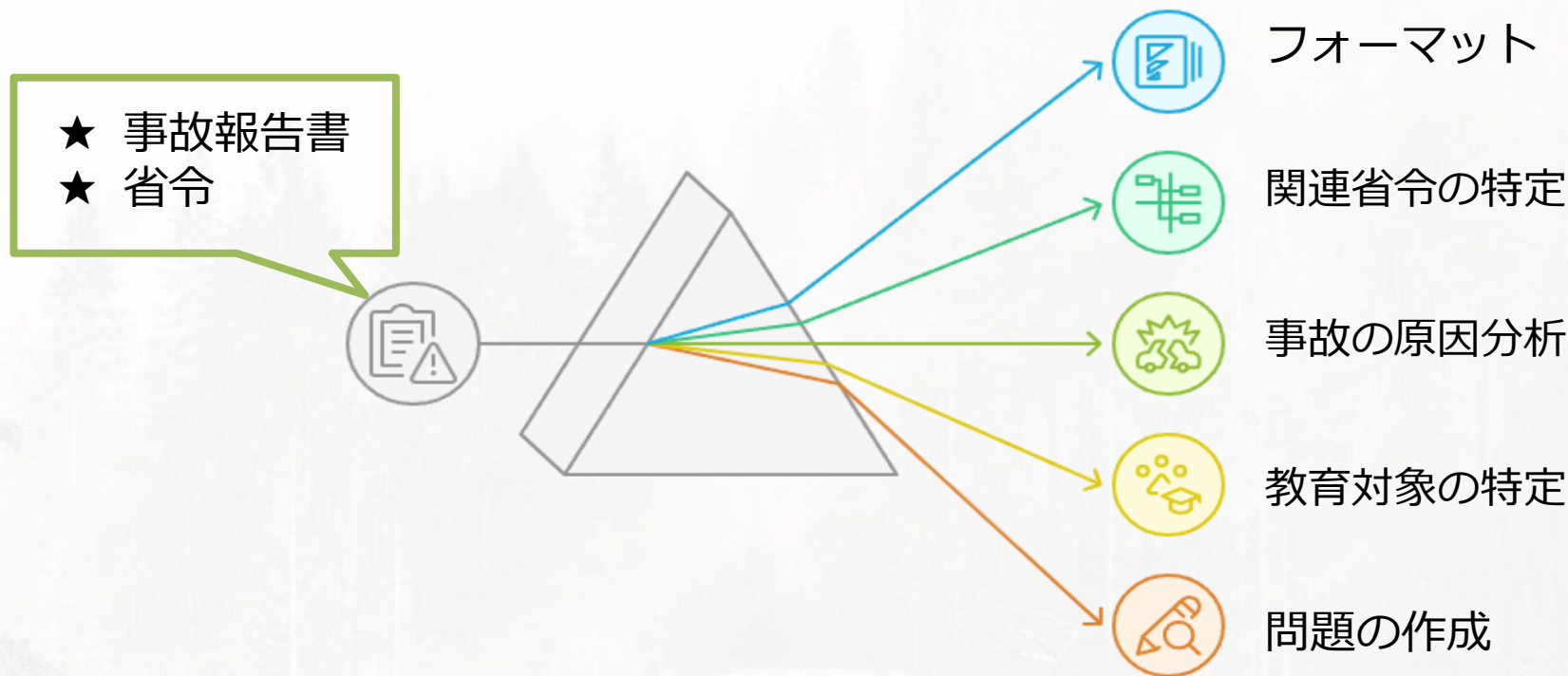
鉄道ナレッジシステムのシステム構成図（前期での開発個所を**赤**で強調）





# 作問 | 技術詳細

事故事例と関連省令から教育対象の特定と教訓・技術的問題を作成



## 作問と教訓抽出における工夫点

- ★ 必要な情報を埋める形に出力フォーマットを事前にプロンプトで定義
- ★ 関連省令箇所を特定しやすいように省令を章ごとに要約
- ★ より深く学べるように関連キーワードをセットで出力

# 作問 | 評価と考察

## 作問結果の妥当性の目視確認の結果

事故を踏まえた作問ができている一方で、模範解答の内容が抽象的である。  
そのため、穴埋問題や短答問題などは良い作問ができている。  
一方で、抽象問題や具体問題では模範解答が不明瞭なため改善が必要。

### 具体的な事故事例（報告書番号: RI2024-1-1）

事例：

列車の走行中にドアの1か所が開いた。

原因：

列車の右側のドア開指令線である315線が走行中に意図せず加圧された際に、戸閉め保安回路が正常に機能しなかったこと。

対象者：

保守作業員

問題：

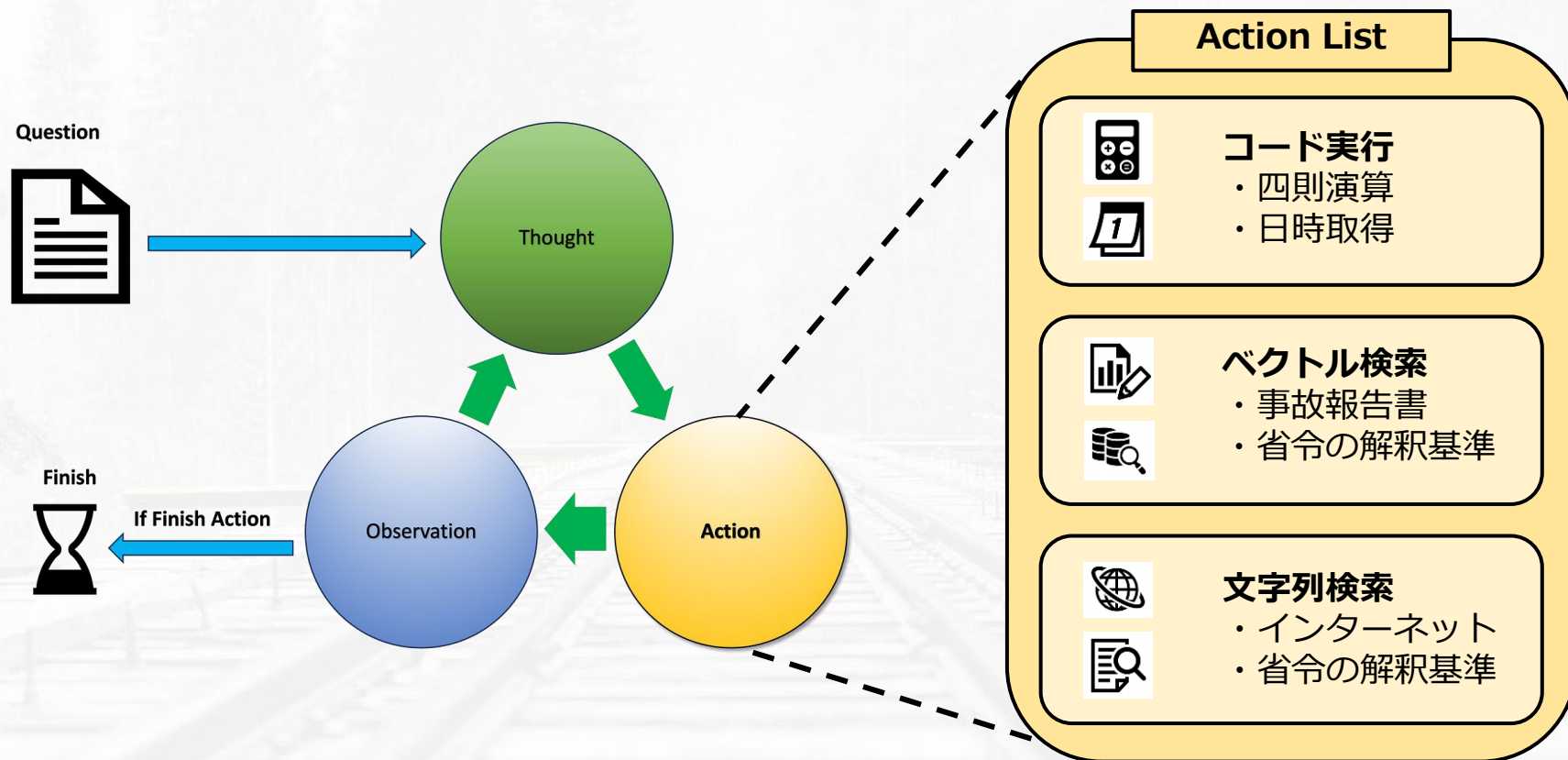
車両装置の保守に関する省令の内容を挙げ、それに基づき確認すべきポイントは何か。



# 回答 | 技術詳細

## ReAct技術の導入により、鉄道技術文書の高度な質問解答精度を実現

前期PBLでは、数値や文書内に類似単語を含む質問への誤答が課題  
そこで、ReActを導入し、言語モデルが思考しながら適切なActionを実行することで、鉄道に関する専門的質問に高精度で答える仕組みを構築



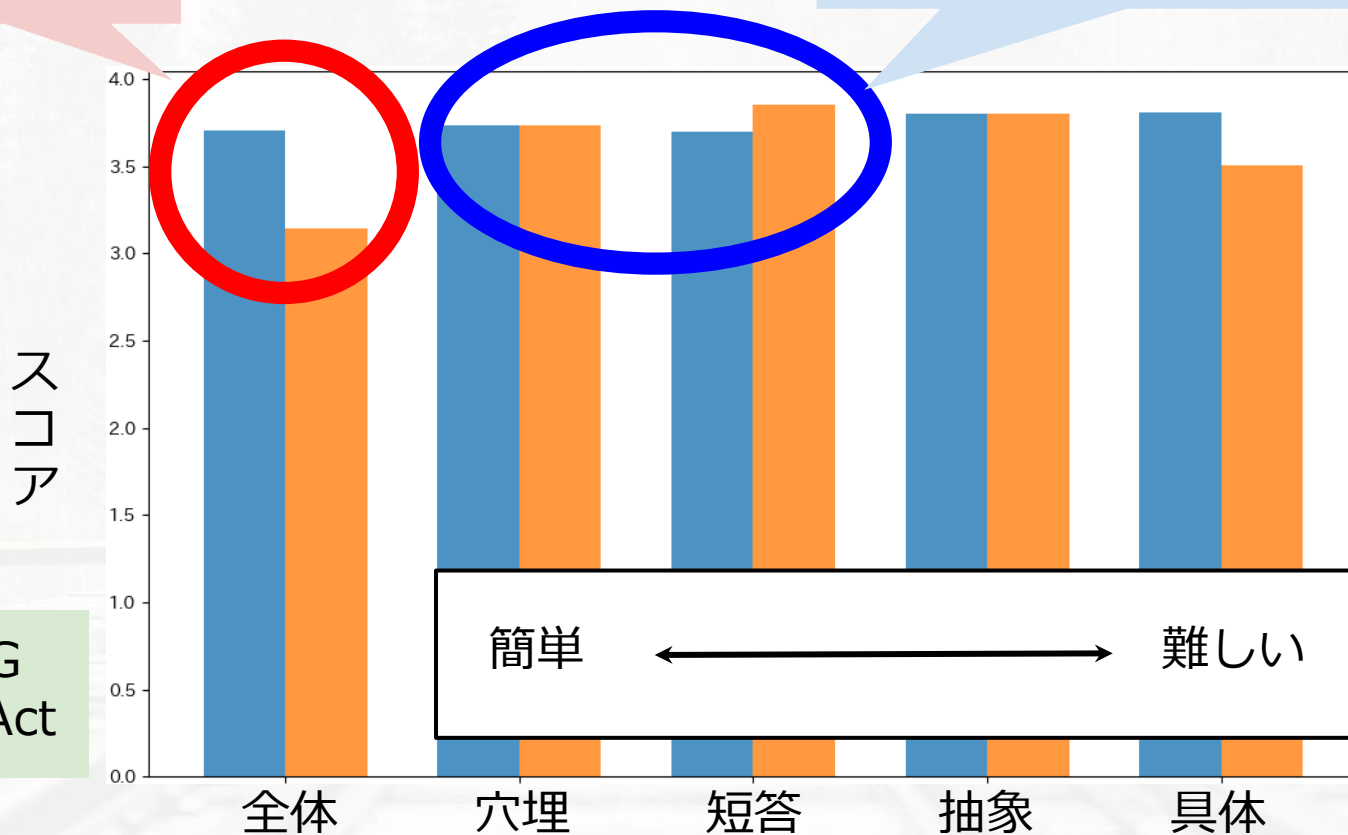
# 回答 | 評価と考察

## ReActの導入により鉄道ナレッジシステムの精度が向上

ReActの導入により全体的な精度は向上

ReActの導入で穴埋問題や短答問題といった簡単なタスクでは精度が悪化。

簡単なタスクではReActの自由度の高さが妨げとなった可能性がある。

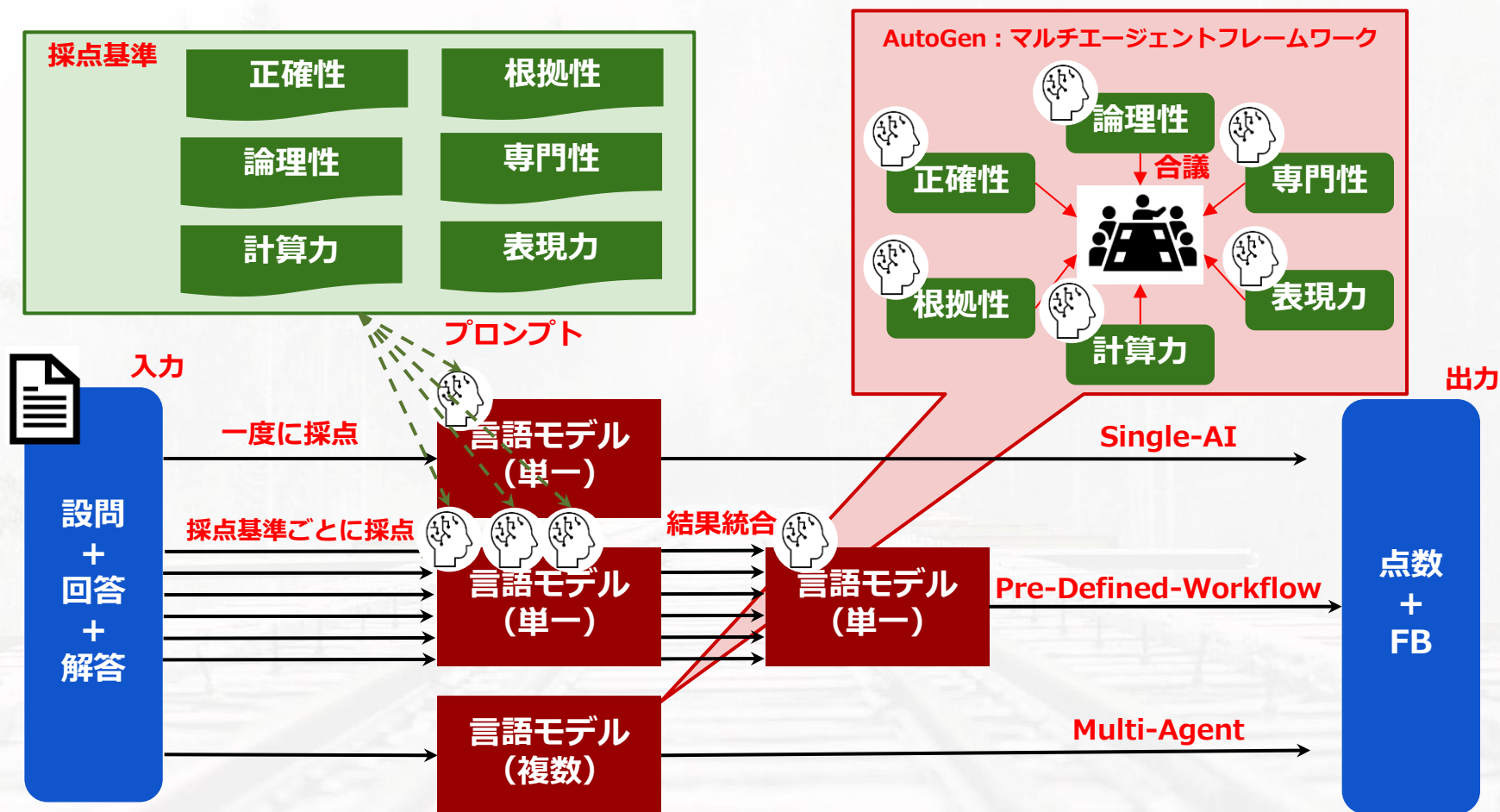


橙色 : RAG  
青色 : ReAct

# 採点 | 技術詳細

## 3種類の採点機構(Single-AI/Pre-Defined-Workflow/Multi-Agent)の比較

6つの評価項目の活用方法として、3種類の採点機構を実装  
手採点と比較した絶対差で3種類の採点機構を評価



# 採点 | 評価と考察

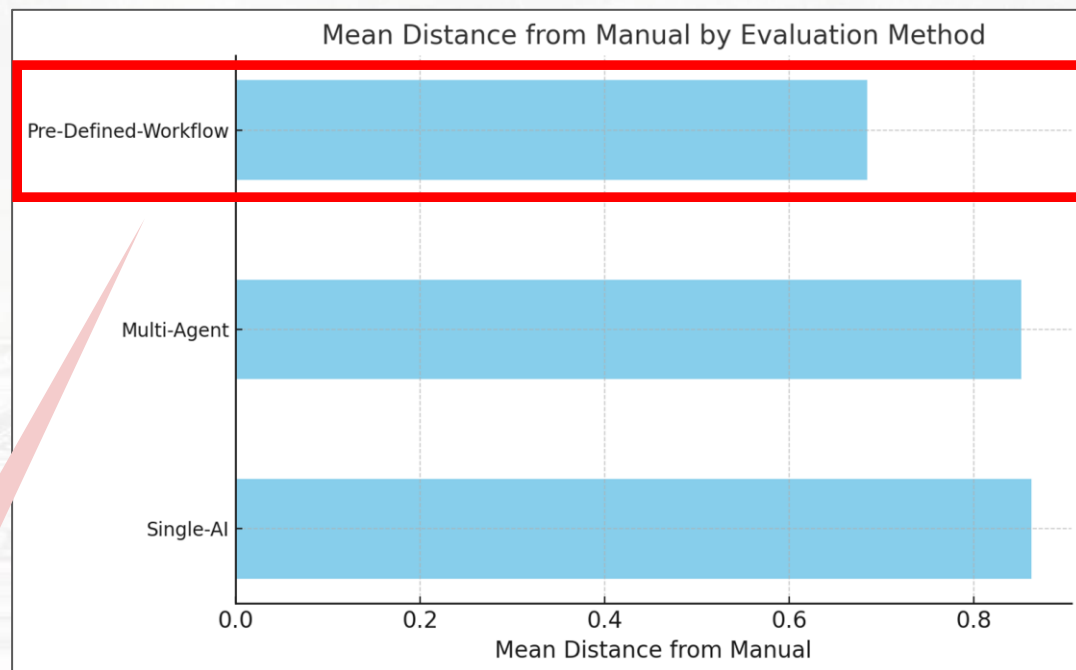
言語モデルによる評価が手採点に近づく鍵は、評価項目の独立性

3種類の採点機構の精度を比較した結果、Pre-Defined-Workflowが最も高い精度を示し、次いでMulti-Agent、Single-AIの順番となった。

Pre-Defined-Workflowが最も優れていた理由として、Pre-Defined-Workflowの採点プロセスの独立性の高さが考えられる。

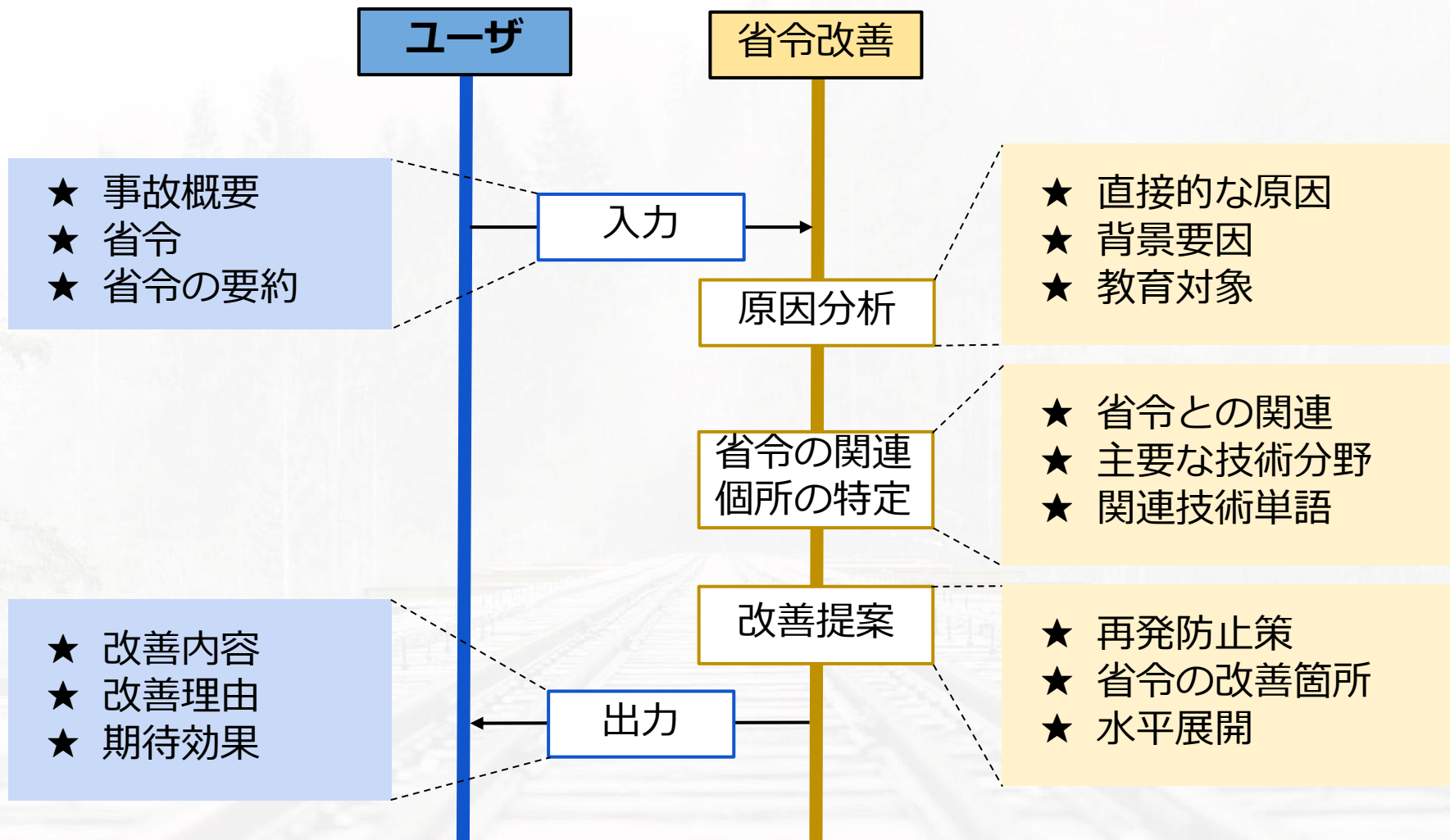
Pre-Defined-Workflowでは各評価項目を独立して言語モデルに入力し、並列的に採点を行う。

これにより、評価項目同士が独立し影響を受けにくく、精度の高い採点が実現できたと考えられる。



# 改善提案 | 技術詳細

事故報告書を入力し、事故の再発防止を目的とした省令の改善案を提案





# 改善提案 | 考察

## 事故の再発防止を目的とした省令の改善提案機能の紹介

改善内容は再教育やハードウェアの増設による改善案は妥当性が高く、システム面の改善提案は具体性が乏しく、必要性のみが指摘されていた。  
また、侵入や飛び込み等の対応の仕組化が難しい事故には改善が難しい。

### 具体的な事件事例（報告書番号: RI2024-2-1）

事例：

運転士が連絡電話を使用中、停車中の車両が逸走し約31m移動。  
車両には21名の乗客が乗車していたが、負傷者はなし。

原因：

降車した際、外とうがブレーキハンドルと接触し、ハンドルが「重なり」位置から「緩ゆるめ」位置に移動。

改善提案：

運転士の服装規定を見直し、安全基準を策定。  
ブレーキハンドルに保護カバーを取り付け、誤触を防止。

# 追加取り組み事項

時間の都合で割愛したが、その他取り組んだことの紹介

## 鉄道技術用語抽出

鉄道技術用語抽出では、TF-IDFやBM25などの機械的なキーワード抽出手法を用い、省令から鉄道に関わる技術用語を抽出することを目的に実施した。



鉄道技術用語のワードクラウド

## AIサイエンティスト

Sakana.aiの提供するAI-Scientistを用いて鉄道業務支援向け言語モデルの研究開発を行った。

アイデア出しにおいては、当班のブレストでも出た「対話型訓練」「AI意見箱」「学習補助」などのアイデアがあり、興味深い結果

一方で、実験においては、有効な成果は得られず、数値評価可能な内容向きの印象

# まとめ

前期から、ReActの導入による回答精度向上や、作問・採点の自動化、  
事故報告から省令の改善提案機能を加え、鉄道ナレッジシステムを拡充

