

# メディアの高品質な知識資産を活用したRAGデータベースの構築と 利用実績等に基づいたフェアな対価還元エコシステムの構築に向けた調査・実証

<b>実施者</b>	note株式会社
<b>概要</b>	<p>生成AIでの利活用のために、メディア等のデータホルダーが保有する高品質な日本語データを集約し、RAGデータベースを構築する。そしてRAGの利用実績等に基づき権利者へ公正な対価を還元する。</p> <ol style="list-style-type: none"> <li>① 出版社・学術機関等から高品質な知識資産を収集。権利関係がクリアな状態でRAG利用ができるような共通契約フレームおよび預託システムを構築。</li> <li>② RAG利用の効率性/精度を向上させるための、検索性・信頼性を高めるデータの正規化やメタデータ標準化を行う。</li> <li>③ 権利処理済みのデータを格納し、AI参照量や利用実績を正確にトラッキング可能なRAGデータベース基盤を実装する。</li> <li>④ AI参照量に応じたライセンス収入分配モデルを検証し、データホルダーへの対価還元と新たな収益源を確立する。</li> <li>⑤ 金融・医療等を中心とした生成AIの精度や参照元が重要な領域でユースケースを創出し、本事業の有効性と事業性を実証する。</li> </ol>

## 提案の背景・社会的現状

日本における知的資産の活用が重要

- 日本語の知識資産の利活用は不十分
- 日本の文化や価値観が国際社会から不可視化され、影響力を失う危険
- 出版社や報道機関が保有する一次情報や専門的コンテンツは代替不可能であり、正規流通の仕組み整備が不可欠

知的資産の活用に伴う課題が存在

- 買い切りを求める開発者と、学習寄与の不透明さの中で対価を求めるデータホルダーの利害が一致しない
- 個別契約・権利確認の負担が大きく、特に中小データホルダーの参画が難しい
- 無断利用や不透明な活用が訴訟や取り組み全体への信頼性を損ねている

## 実施内容

- ① 多様な知識資産の収集と、安心して参加できる預託スキームの構築
- ② 検索性に優れた高品質データを整備するための標準化 (メタデータ/スコアリング等)
- ③ RAGデータベースおよびAPI基盤の構築
- ④ 利用実績等に基づく、納得感と透明性をもたらす利益還元ロジックの検証
- ⑤ RAGデータベースの有効性を示す模範的ユースケースの創出と社会発信

## 社会実装の方法

- **RAGデータ基盤の事業化:**  
本事業で開発するRAGデータ基盤を実サービスに実装し、継続運用
- **標準化仕様案の公開:**  
メタデータや契約フレームを公開・更新し、社会基盤を整備
- **ユースケース発信:**  
実証分野の成果を広報し、本エコシステムの有用性を提示

成果物の公開



データセット  
(一部予定)



モデル  
(一部予定)



レポート  
(公開)

# メディアの高品質な知識資産を活用したRAGデータベースの構築と 利用実績等に基づいたフェアな対価還元エコシステムの構築に向けた調査・実証

## データエコシステム図

■ : ステークホルダー

□ : 実証範囲

→ : データの流れ

補足

