

報道関係者各位

2025年5月13日

FastLabel 株式会社

**FastLabel、ストックマーク株式会社へ日本語特化型 LLM 開発に必要な独自データを作成し提供  
ハルシネーションを抑止したドキュメント読解基盤モデルにおける GENIAC 第 2 期を支援  
～ 本発表に関連した事例ウェビナーを 6 月 4 日（水）12 時に開催 ～**

Data-centric AI<sup>[※1]</sup>開発に必要なデータ収集・生成からアノテーション、モデル開発、DataOps 構築までの全工程を支援する FastLabel 株式会社（本社：東京都新宿区、代表取締役 CEO：鈴木健史、以下「当社」）は、ストックマーク株式会社（本社：東京都港区、社長：林 達、以下「ストックマーク」）へ、ハルシネーションを抑止したドキュメント読解基盤モデルの開発に必要な、独自のデータを作成し提供したことをお知らせします。

[※1] Data-centric AI：AI のデータセットを改善することでモデルの精度改善を実現するアプローチ

## Stockmark × FastLabel

今回の当社のデータ作成と提供は、経済産業省・NEDO が推進する国内生成 AI 開発プロジェクト（以下、GENIAC）<sup>[※2]</sup> の第 2 期（2024 年 10 月発表<sup>[※3]</sup>）の中で行われ、ストックマークによる高い日本語能力と深いビジネス知識を併せ持ったビジネスドメインで信頼して活用できる複雑なドキュメント読解のための LLM 開発に活用されました。

[※2] GENIAC：経済産業省が 2024 年 2 月に立ち上げた国内の生成 AI 開発力強化プロジェクト

[※3] スtockマークの発表：<https://stockmark.co.jp/news/20241010>

### 【日本の LLM 開発における課題】

日本の LLM 開発は主にデータと人員においていくつかの課題があります。

**データの課題：**日本語は英語圏に比べて公開データが少なく、特に高品質な公開データセット（書籍、論文、SNS、ニュース記事など）が限られていて、ライセンス面でも制約が多いです。また、日本語特有の表記揺れや曖昧さも多く、同じ言葉でも漢字・ひらがな・カタカナ・ローマ字の混在、文脈依存の意味変化や助詞・敬語表現の複雑さなど、モデルが学習しづらい要素が多いです。

**人員の課題：**日本国内において LLM 開発をリードできる AI エンジニア・研究者は限られており、実際に大規模なモデルの学習・運用を行った経験者が少ない中で開発を推進していかなければなりません。そのため、本質的な研究や開発の業務ではないデータ作成や収集に、限られた人員のリソースが割かれてしまうことは本意ではありません。

### 【当社の支援内容】

課題を踏まえ、今回当社ではストックマークへ以下の支援を行いました。

**AI 開発に必要なデータの作成：**ストックマークが指定する数万枚のパワーポイントデータをテキストデータに変換しつつ、資料内に含まれる図版やグラフなどの意味や意図の読み取りを含めて実施。

**GENIAC 要件への対応：**GENIAC における当社の豊富な採択支援実績に基づき、審査基準を満たすデータ作成を支援。

本プレスリリースに関するお問い合わせ

FastLabel 株式会社 広報担当 [pr@fastlabel.ai](mailto:pr@fastlabel.ai)

お問い合わせフォーム URL <https://fastlabel.ai/contact-pr>

**【ストックマーク Senior Researcher 森長様より本支援に関する所感】**

この度は、FastLabel 様へパワーポイントデータへの詳細なアノテーションデータの作成を依頼しました。本プロジェクトに際しては図や表、画像といったテキストで表現されていないデータの意味をどうモデルに学習させるかといった課題がありましたが、アノテーションにあたり、図や表、画像の意味や意図を可能な限り自然言語でテキストとして変換していただいたことで円滑に開発を進行できたと感じています。

**【本プレスリリースに関連した事例ウェビナーのご案内】**

開催日時：2025年6月4日（水）13時00分～14時00分

タイトル：GENIAC 採択プロジェクトに学ぶ！LLM 開発の  
現在地と課題解決のリアル

概 要：GENIAC の採択企業 2 社による研究テーマや成果、  
課題への具体的なアプローチを紹介

登 壇 者：ストックマーク株式会社、NABLAS 株式会社、  
FastLabel 株式会社

参 加 費：無料

申し込み：<https://fastlabel.ai/seminar/lp-20250604>

**【当社事業について】**

当社は「データセット提供」、「アノテーション代行」、「モデル開発支援」、「FastLabel Data Factory」など、AI 開発を行うお客様に向けたトータルソリューションを提供しています。

データセット提供：権利クリアかつ高品質な各種データの収集、販売を行います。ストックの提供や新規撮り下ろしにも対応しています。

アノテーション代行：あらゆる非構造化データに対応し、弊社独自の品質管理によりデリバリー品質を担保しています。ドメイン知識が必要な仕様にも対応可能です。

モデル開発支援：画像や動画の撮影条件や正確性／統一性を読み解きながらモデルの学習・評価を実施し、評価結果を精度向上につなげていきます。

FastLabel Data Factory：データ収集・管理、アノテーション、モデル開発までワンストップで提供する SaaS です。DataOps 構築を実現し教師データ準備を大幅に効率化できます。

当社は各業界リーディング企業への豊富な支援実績を有し、権利クリアかつ高品質なデータ作成に強みを置いております。豊富な経験を通して培ったアノテーション仕様作成をはじめとした AI 開発のノウハウで、今後も多くのお客様を支援してまいります。

**【当社の概要について】**

社 名：FastLabel 株式会社

代 表 者：代表取締役 CEO 鈴木健史

事 業 内 容：Data-centric AI 開発を支援するプロフェッショナルサービスとプロダクトの提供

設 立：2020年1月23日

本 社 所 在 地：〒163-0224 東京都新宿区西新宿 2-6-1 新宿住友ビル 24 階

U R L：<https://fastlabel.ai/>

本プレスリリースに関するお問い合わせ

FastLabel 株式会社 広報担当 [pr@fastlabel.ai](mailto:pr@fastlabel.ai)

お問い合わせフォーム URL <https://fastlabel.ai/contact-pr>