

2024 年 11 月 28 日

東洋経済新報社と Visual Bank

生成 AI 向け学習用データに関する
データパートナーシップ契約を締結

東洋経済の保有するテキストデータ等を『Qlean Dataset』において、
LLM／生成 AI／マルチモーダル AI 向けの AI 学習用データとして提供開始

 **Qlean Dataset**
by VISUAL BANK**東洋経済新報社**

株式会社東洋経済新報社（東京都中央区、代表取締役社長：田北 浩章、以下「東洋経済新報社」）は Visual Bank 株式会社（東京都渋谷区、代表取締役 CEO：永井 真之、以下「Visual Bank」）傘下の株式会社アマナイメージズ（以下「アマナイメージズ」）を通じて、データパートナーシップ契約を 2024 年 11 月 19 日に締結したことをお知らせします。

東洋経済新報社の有するテキスト等の大規模データを、Visual Bank の AI 学習用データセット開発サービス『Qlean Dataset（キュリンデータセット）』を通じ、AI 学習用途に対して販売開始します。

本契約締結により東洋経済新報社の有する膨大なテキスト、グラフ等のデータを日本の AI 開発現場に届け、日本の AI 開発が加速・発展することに貢献していきます。



提供可能なデータ

『週刊東洋経済』
「東洋経済オンライン」
「週刊東洋経済 e ビジネス新書」
「会社四季報オンライン」
『株式ウイークリー』で
公開・提供される記事、グラフ、
表のデータ

背景

●AI の隆盛と良質な日本語コーパスの必要性

人工知能（AI）技術の進化のスピードは著しく、社会・生活の様々な場面に活用されています。特に 2022 年以降は生成 AI やマルチモーダル AI の技術革新が注目を集めています。LLM（大規模言語モデル）のハルシネーション 軽減を含めた自然言語処理技術の向上において、豊かかつ良質な学習用データセットは不可欠です。しかし、グローバルで英語を用いた LLM の研究開発が進む一方で、日本においては信頼性の高いテキストデータを大量に入手する手段が限られており、日本語の独自 LLM の開発は依然としてデータの質と量がボトルネックになっています。代表的な日本語公開コーパスとしては、CC-100、mC4、OSCAR、Swallow Corpus 等がありますが、品質のばらつきもあり、学習データとして利用するためにはデータのクリーニングなどの前処理が必要になります。

●テキスト、グラフ等の大規模データ提供による AI 開発現場の支援を目指したデータパートナーシップ契約

こうした背景の中、東洋経済新報社と Visual Bank は、AI の研究開発におけるデータの多様性と品質を確保し、イノベーションを加速させることを目的とし、日本市場に向けた AI 学習用のテキスト、グラフ等の大規模データの提供を開始します。LLM・生成 AI・マルチモーダル AI での活用を念頭に、良質で信頼性の高い日本語コーパス等を提供することで、日本における AI の研究開発を支援していきます。

データパートナーシップ契約の内容

本契約は、東洋経済新報社の有する良質なクローズドなテキスト、グラフ等のデータを、Visual BankのAI学習用データセット開発サービス『Qlean Dataset (キュリデータセット)』を通じて販売開始するものです。AI開発企業や研究機関に対して提供し、LLM/生成AI/マルチモーダルAIの学習用途にご利用いただくことを想定しています。

提供可能なデータ

「東洋経済オンライン」「会社四季報オンライン」『株式ウイークリー』『週刊東洋経済eビジネス新書』『週刊東洋経済』で公開・提供される記事、グラフ、表のデータが対象です。東洋経済新報社の有する、ビジネス誌系サイトのトップランナーでもある「東洋経済オンライン」や株式投資領域で独自のポジションを築いている「会社四季報オンライン」をはじめとした、過去20年以上のテキストデータやグラフデータを提供可能。各業界の専門記者が執筆してきた良質な日本語データを提供します。



提供するデータの形式

本契約の対象となるデータは、以下の形式で提供されます。
1. 記事本文データ
2. グラフ・表データ
3. テキストデータ
4. テーブルデータ
5. 図表データ
6. 表データ



記事テキストデータ

グラフ・表データ



Qlean Dataset (キュリンデータセット) とは

AI 開発の加速を支援するために、高品質かつ権利クリアなデータセットを開発するサービスで、40 年にわたり権利許諾のノウハウを持つアマナイイメージズが提供。画像・映像・テキスト・音声・3D など多領域の「権利許諾済みデータベース」を基盤にデータセットを開発し、利用に応じた対価をデータ権利者に還元。本契約では、データセットの企画・開発と販売機能を提供。

<https://qleandataset.amanaimages.com/>

「東洋経済オンライン」及び「会社四季報オンライン」を含む東洋経済新報社が提供・公開している記事・テキストデータベースは、東洋経済新報社の著作物で、日本の著作権法や国際条約などで保護されており、かつ、『Qlean Dataset』を通じて AI の学習用に有償で提供しているデータとなります。スクレイピングやクローリングなどいずれの方法を問わず、無断での AI の学習利用を禁止致します。万が一、無断利用が発覚した際は、著作権の侵害として法的措置を含め厳正に対処します。

各社コメント

Visual Bank グループ 株式会社アマナイイメージズ 代表取締役 CEO：望月 逸平様

日本において生成 AI・LLM の研究開発が急ピッチで進行しています。そのなかで、良質なテキストやグラフデータを求める声は多く聞かれます。東洋経済新報社様とのデータパートナーシップ契約により、東洋経済新報社様の有する貴重な素材を『Qlean Dataset』を通じて AI 開発現場に届けることで、新たなテクノロジーやユースケースが創出される一助になることを目指します。Visual Bank グループは、今後も適切な権利処理・用途管理・対価還元のもと、クローズド・データの AI 学習用としての社会での利活用を推進し、安心・安全な AI 開発を裏側から支援していきます。

『Qlean Dataset』：<https://qleandataset.amanaimages.com/>

株式会社東洋経済新報社 取締役：廣田 充彦

この度は Visual Bank 様にお力添えをいただき、「東洋経済オンライン」や『週刊東洋経済』に掲載している記事などをご提供させていただくこととなり大変有り難く思います。AI 規制が進むなかで、法的に問題がないことが明らかな学習用データを提供している『Qlean Dataset』の存在が重要性を増しており、その充実に微力ながら貢献して参ります。

【Qlean Dataset × 東洋経済新報社のサービス概要】

特設 LP はこちらをご覧ください。

<https://qleandataset.amanaimages.com/toyokeizai>

●特長

- ①東洋経済新報社の過去 20 年以上の記事テキスト、グラフデータの提供
- ②LLM/生成 AI/マルチモーダル AI の学習用データに使用可能
- ③専門的なインストラクションデータも作成可能

【LLM（大規模言語モデル）/生成 AI/マルチモーダル AI における想定ユースケース】**●LLM の開発**

高品質な日本語コーパスを利用した汎用性の高い LLM 開発、ファインチューニング
経済領域での特化型の LLM 開発、ファインチューニング

●図表×テキストのマルチモーダル AI の開発

IR 資料、決算資料に関連するマルチモーダル AI

社内パワーポイント資料、エクセル、ワード、PDF などの資料に関連するマルチモーダル AI

お問い合わせ先

【AI 開発ベンダーの方からのお問い合わせ】

Visual Bank 株式会社 Qlean Dataset お問い合わせフォーム

<https://qleandataset.amanaimages.com/contact>

本リリースに関するお問い合わせ先

株式会社東洋経済新報社 総務局広報室：青柳・遠藤・江口

TEL：03-3246-5404 FAX：03-3279-0332 email:org-Somu-Koho-Shitsu@toyokeizai.co.jp