



【Python】AI学習用コーパス（テキストデータ）の自動検品・検証プログラム開発

完全在宅（フリーランス）でのプログラミング業務

Job Information

Hiring Company

Franchir Co., Ltd.

Job ID

1609982

Industry

Software

Company Type

Small/Medium Company (300 employees or less)

Non-Japanese Ratio

Majority Japanese

Job Type

Freelance

Location

Tokyo - 23 Wards, Toshima-ku

Train Description

Yamanote Line Station

Salary

Negotiable, based on experience

Hourly Rate

出来高払い

Refreshed

September 29th, 2026 16:00

General Requirements

Minimum Experience Level

Over 1 year

Career Level

Mid Career

Minimum English Level

Business Level (Amount Used: English usage about 50%)

Minimum Japanese Level

Fluent

Minimum Education Level

Bachelor's Degree

Visa Status

Permission to work in Japan required

Job Description

AI（大規模言語モデルなど）の学習に使用する「コーパスデータ（主にJSON / CSV / 正解ラベル付きテキスト）」の品質を自動でチェックするスクリプト・プログラムの開発をお願いします。

【具体的なチェック項目】

- データのフォーマット・構造チェック（ネストの崩れ、文字コード、空データの検出）
- テキストの基本バリデーション（禁止文字の混入、文字数制限の超過など）
- アノテーション（タグ付け）の整合性チェック（開始タグと終了タグの不一致など）
- ※要件定義から一緒に相談しながら進めさせていただける方を歓迎します。

Required Skills**【必須スキル・経験（Must）】**

- **Python**を用いたデータ処理、テキスト処理の実務経験（2〜3年以上）
- **正規表現（RegEx）**を用いた高度な文字列パターンの抽出・照合スキル
- **pandas**などのデータ解析ライブラリの使用経験
- Git / GitHub を用いたソースコード管理

【歓迎スキル・経験（Want）】

- **自然言語処理（NLP）**に関する知見、ライブラリ（GiNZA, Spacy, MeCabなど）の使用経験
- AIコーパス、アノテーションデータの取り扱い経験
- 大量データ（数万〜数百万件）を高速に処理するプログラムの最適化経験

【求める人物像】

- 単なる指示通りではなく、データの不備を見抜くための仕様を一緒に考えてくれる方
- 細かい条件分岐や例外処理をロジカルに組み立てられる方
- 報告・連絡・相談がスムーズで、納期を守れる方

Company Description