



CREATIVE CONNECTIONS
& COMMONS INC.



多言語AIデータサービス

カスタム音声・テキスト・会話データセット
大規模プロジェクトにも対応

CCCについて



CCCは、30以上の言語に対応する翻訳・ローカライゼーションの専門企業です。こうした言語的な専門性を活かし、タグリッシュ（タガログ語と英語が混在するフィリピン特有の話し方）をはじめ、文化的ニュアンスやコードスイッチングを含む多様な言語表現にも対応。お客様のユースケース、ワークフロー、プロジェクト規模に合わせて、AI向けの音声・文字起こし・翻訳データセットをカスタム構築します。



CCCの目指すもの

翻訳・ローカライゼーション、多言語コミュニケーションで培ったCCCの専門知識を、高品質なオーダーメイドのデータセットへと活かすことで、人間の言語とAIをつなぐ架け橋となることを目指しています。それにより、AIシステムが多様な言語、十分なデータが整備されていない言語、さまざまな文化、そして現実世界の文脈を、より自然かつ正確に理解できるよう支援します。

CCCの言語データサービス



🎧 音声収録サービス

AIの事前学習・ファインチューニング向けに、台本収録、会話形式、会議形式など、さまざまな音声データを提供します。ポッドキャスト形式の音声をはじめ、Zoomを利用したセッション、オンライン会議、スタジオ収録などに対応しています。ファインチューニング向けには、即興形式・台本形式のいずれにも対応し、高品質な48kHz音声の収録も可能です。

🗣️ 音声データの文字起こし・品質評価

既存の音声データセットやAI生成の文字起こしに対し、人手による文字起こし、評価、ポストエディティングを行います。これにより、自動システムでは見落としや誤認識が起りやすいスラング、専門用語、コードスイッチング、コードミキシング、アクセント、文脈依存の表現なども、正確に捉えることができます。また、話者情報、感情、ノイズなど、各種メタデータのアノテーションにも対応しています。

📄 会話・特定シーン向けテキスト生成

メッセージ、メール、カスタマーサポートのシナリオ、対話、音声収録用の台本など、用途に応じたテキストデータセットを人手で作成します。自然な言語表現や文化的ニュアンスを反映した、信頼性の高い正解データ（グラウンドトゥールース）が必要なユースケースに適しています。これらのデータは、特に多言語環境や十分なデータが整備されていない言語、コードスイッチングを含むデータ、AIモデルの評価、そして合成データだけでは実際のコミュニケーションを十分に再現できないユースケースに有用です。

🔧 機械翻訳のポストエディティングおよび評価

機械翻訳やLLM（大規模言語モデル）が生成した多言語コンテンツに対し、人手によるポストエディティングと評価を行い、正確性、流暢さ、用語の適切性、文化的な適合性、指示への準拠状況を検証します。CCCの言語専門家は、選好評価、エラーアノテーション、品質評価、人手による修正に対応し、AIモデルの評価・改善に役立つ信頼性の高いフィードバックデータを提供します。

対応言語



★ アジア太平洋地域

日本語
韓国語
タガログ語
セブアノ語
タイ語
マレー語（マレーシア）
インドネシア語
中国語（繁体字）
中国語（簡体字）
クメール語
ネパール語
モンゴル語

ベトナム語
ビルマ語
ヒンディー語
タミル語
ウルドゥー語
ベンガル語
★ ヨーロッパ
スペイン語（スペイン）
ポルトガル語（ポルトガル）
イタリア語
フランス語
ドイツ語
ウクライナ語

ロシア語
ギリシャ語
ベラルーシ語
スロバキア語
ポーランド語
ハンガリー語
スロベニア語
チェコ語
英語（英国）
★ 南北アメリカ
スペイン語（ラテンアメリカ）
ポルトガル語（ブラジル）
英語（米国）

★ 中東・コーカサス・アフリカ

アルメニア語
ジョージア語
ペルシア語
トルコ語
アラビア語
アフリカーンス語
スワヒリ語
キニャルワンダ語
ズールー語

ご希望の言語が一覧にない場合も、お気軽にお問い合わせください。

プロジェクトごとの柔軟な価格設定



パイロットデータセットから大規模な多言語プロジェクトまで、お客様のご要件に合わせて、対応範囲と価格を柔軟にご提案します。

対応内容は、**対象言語**、**データ量**、**要件の複雑さ**、収録仕様、アノテーション要件、納期などに応じて個別に設計します。プロジェクト要件に合わせたご提案・お見積もりをご用意します。まずはお気軽にお問い合わせください。



お問い合わせはこちら

www.cccjph.com | inquiry@cccjph.com | (082) 233 0032

新たなストーリーを創る

