



AIによるコンテンツ最適化と生成の ケーススタディ 生成系 AI を利用したコンテンツ最適化と ローカリゼーションの実験

4つのAIツールでテスト

2つの言語でプロンプトを記述してテスト

英語のプロンプトで最適化したドイツ語コンテンツの平均品質スコア: 86%

今回の実験の目的

今回の実験は、ライオンブリッジの社内チーム、リソース、コンテンツを使用して行いました。特に、使用したドイツ語のコンテンツは、当社の公開済みのブログ記事から抜粋したものです。本ケーススタディは当社のコンテンツ (特にドイツ語) が中心ですが、実験の目的は、あらゆる業種と言語を対象に、当社のお客さまに活用できるようなプロトコルを作成し、知見を集めることにありました。当社は今回の実験の結果と内容を踏まえ、さまざまなコンテンツ ニーズや言語に対応したサービスをお客さまに提供するとともに、この情報を自社のマーケティング コンテンツの制作プロセスおよび最適化プロセスで利用する予定です。

課題

コンテンツの最適化には、特に英語以外の言語によるコンテンツの場合に多大な時間とコストがかかることがあります。また、既存のコンテンツを翻訳する場合は、SEO の面で必ずしも最良の結果が得られるとは限りません。ときにはコンテンツを単に直訳しただけの場合もあり、言語の持つニュアンスや地域固有のキーワード、ユーザーの行動やデータは反映されません。生成系 AI ツールを使用することで、コスト、時間、手間を抑えながら多言語コンテンツを最適化できる可能性があります。

背景

使用した4つの生成系 AI ツール、Amazon Bedrock、ChatGPT (ChatGPT 4 を使用した Lionbridge Azure インスタンス)、Google Bard、Meta の Llama はいずれも、コンテンツの制作と最適化のための機能を備えています。

これらのツールは、検索パターン、競合他社のデータ、ユーザーの行動を分析し、キーワードとコンテンツについてデータに基づいた提案を行います。

それぞれのツールは、可能な限りユーザーの期待に沿い、魅力的で SEO の効果が高いコンテンツになるように最適化を支援します。場合によっては、人間が作成したものよりも優れたコピーを作成することもあります。

たとえば、厳密な文字数制限に従ってコンテンツの制作や翻訳を行う場合には、(特に SEO の効果の面で) 生成系 AI ツールは人間に匹敵する能力を発揮する場合があります。

解決策

当社では、特に英語以外の言語における調査やコンテンツの生成に関してどのツールが最も効果的かを確認するために、それぞれの生成系 AI ツールを使用して3フェーズ構成の実験を行いました。今回の調査では、多言語コンテンツにおけるプロンプト作成のベストプラクティスを見出すことに力を入れました。また、古いコンテンツを新しい言語で最適化して Web サイトの SEO 効果を高めるという状況を想定し、その場合のベストプラクティスを探りました。

ライオンブリッジのアプローチ

今回の実験では、Amazon Bedrock、ChatGPT (ChatGPT 4 を使用した Lionbridge Azure インスタンス)、Google Bard、Meta の Llama の 4 つの AI ツールを使用し、調査、制作、分析の3つのフェーズを2回実施しました。

調査: このフェーズでは、キーワードの調査を行い、ドイツ語でコンテンツを制作する場合に最も効果的な2つの質問を選ぶように各 AI ツールに指示しました。

それぞれのツールには次の内容を入力しました。

- ペルソナ プロンプト: ドイツ語のコンテンツライターとして振る舞い、「新しいドイツ語のキーワード3つとドイツ語の質問3つを出力する」ように指示
- 最適化するソースコンテンツ
- 既存のキーワード

加えて、後に短い段落から成るコンテンツを生成するため使用する2つの補足質問を選択しました。

キーワードの調査においては、人間参加型の AI の活用 (一般に「ヒューマンインザループ」と呼ばれます) が不可欠であると思われます。特に Bard でその傾向が顕著でした。Bard は調査では優れた成果を上げましたが、出力した選択肢が多すぎたため、当該分野の専門家がコンテンツの生成に必要なキーワードや質問を選択する必要がありました。



制作: このフェーズでは、ドイツ語で2組のプロンプトを作成し、それらのコンテンツ制作プロンプトに前のフェーズで特定したキーワードを含めました。1組のプロンプトはコンテンツライターのペルソナに基づいたもの、もう1組はタスクの指示です。

今回の実験では、すべてのプロンプトで最大限の効果を発揮できるように、当社AI部門のメンバーにレビューと調整を依頼しました。最終的なプロンプトの決定後に(後述の「プロンプト」セクションに当該プロンプトを示します)、当社の翻訳者にこのプロンプトをドイツ語にローカライズしてもらいました。その後、それら2組のドイツ語プロンプトを4つのAIツールに入力し、提案されたコンテンツの収集とスコア判定を行いました。

プロンプトの作成には以下の入力を使用しました。

- 調査フェーズで特定した質問とキーワード (ドイツ語)
- 手本にさせたい良質なコンテンツのサンプル (ドイツ語)
- ソース記事のテキスト (ドイツ語)
- ペルソナ プロンプト (英語およびドイツ語)
- タスク プロンプト (英語およびドイツ語)

プロンプトは2組あるため、AI ツールごとにプロジェクトを2回ずつ実行する必要がありました。

分析: このフェーズでは、ドイツ語を話すビジネスと言語の専門家10名にコンテンツのレビューを依頼しました。先入観を避ける「ブラインドレビュー」を行うために、どのツールから生成されたコンテンツかは伝えませんでした。

評価対象の出力は、以下の2つのスコア判定方法を組み込んだスプレッドシートに記しました。

- 「使用できない」、「編集すれば使用可能」、「編集なしで使用可能」の3つの値を選択できるドロップダウンメニュー
- コンテンツを1～100でスコア判定するためのフィールド

また、フィードバックを収集してレビューを行い、以下を判断しました。

- どのツールが最も効果的だったか
- プロンプト作成のベストプラクティスは何か、他の言語によるプロンプト作成についてはどうか
- 各ツールでキーワード調査をどの程度適切に行うことができるか
- 各ツールでコンテンツに対する質問をどの程度適切に生成できるか
- 各ツールで記事に対する観点をどの程度適切に生成できるか
- 生成された質問に対する回答を各ツールでどの程度適切に作成できるか
- AIによる最適なコンテンツ制作プロセスを実現するにはどうすればよいか
- 効果的なプロンプトを作成するにはどうすればよいか

プランニング



使用する生成系AIツールを3つ選択



実験で最適化する記事を3つ選択



レビューを行う専門家10名を決定



キーワードと質問を調査



プロンプトを作成してドイツ語にローカライズ

使用したプロンプトの例

ペルソナプロンプト

- あなたはライオンブリッジのコンテンツライターで、言語サービスに関するコンテンツを制作しています。
- あなたはドイツ語を話し、ドイツに住んでいます。
- あなたは、ドイツ、オーストリア、スイスなどにいるドイツ語を話す人を対象にドイツ語のコンテンツを制作しています。

オーディエンスプロンプト

- マーケティング部門またはコミュニケーション部門に勤務する人々
- シニア マネージャー、ディレクター、副社長、社長、CMO、CEO など、上級職に就いている人々
- ドイツ、オーストリア、スイスなどに住み、ドイツ語を話す人々

制作プロンプト

- この記事 ... に追加する 2 段落の文章を記述してください。
- それぞれの段落には、これらの質問 ... に対する一つの回答を含めてください。
- この URL のプライマリキーワードは ... です。
- セカンダリキーワードは ... です。

複数の部門から編成されたチーム

- ドイツ語を話すプロンプトエンジニア 10 名
- ライオンブリッジの SEO 専門家 1 名
- ライオンブリッジのデジタルマーケティング専門家 1 名



結果

AI は、あらゆる最適化プロジェクトで非常に重要な要素になると思われます。当社の実験では、AI ツールが既存のコンテンツ (英語以外のコンテンツも含む) を効果的かつ効率的に最適化して更新するのに役立つ可能性があることが確認されました。

前向きな結果として、実験の進展に伴って出力品質が向上しました。コンテンツの最適化や制作のあらゆるプロセスにおいては、ヒューマンインザループによる人間参加型の AI 活用が非常に重要になります。AI の出力から最適化のためのキーワードを選択するには人の介入が欠かせず、また出力された最終的なコンテンツを検証するレビュー担当者も必要です (ただし、出力品質には引き続き大幅な向上が見られるため、将来的にこの状況は変わる可能性があります)。

今回特定されたベストプラクティスをいくつかご紹介します。

- 現時点では、ヒューマンインザループで最終的な AI の出力を検証する必要がある。
- 実験プロセスを通じ、新しいベストプラクティスの発見に伴って AI への入力品質が向上した。
- AI では、以下の形式で指示した場合に高品質のコンテンツを生成できる可能性が高い。
 - 短い文章
 - 分割しやすいコンテンツ (上位 10 項目のリストなど)
- 現時点では、AI とのやり取りは英語で行うのが最も効果的である (各ツールは英語でトレーニングされているため、当然の結果と言えます)。ドイツ語のプロンプトを使用しても、AI ツールが生成するドイツ語コンテンツの質は向上しませんでした。実際には、英語のプロンプトを使用した方がドイツ語コンテンツの質が向上しました。注: ただし、ソースコンテンツ、キーワード、質問、例を (英語ではない) ターゲット言語で入力するのは効果的でした。
- コンテンツ制作プロセスでは、コンテンツ制作者と対象オーディエンス層について詳細に記述する必要がある。また、プロンプトライブラリを作成することが理想的です。プロンプトライブラリによってユーザーはプロンプトを再利用し、顧客に対して一貫性のあるコンテンツを制作できます。

調査に関しては、Bard の出力が最も優れていたと考えられます。

Bard の出力は、調査を目的として開発された MarketMuse などのツールが生成した調査の結果と似ていました。調査に関する Bard の出力で注意する点として、コンテンツ制作フェーズに使用する情報を選択する際に人の介入が必要でした。ドイツ語の指示を使用した場合、Bard は 2 つのフェーズで安定した結果を見せましたが、英語の指示による出力では、スコアは 82/100 から 66/100 に変化しました。今回の実験では使用したデータセットが小規模であったため、Bard のスコアの平均 (英語の場合) は 74/100 となりました。

ChatGPT については、コンテンツに対して否定的なコメントがやや多く寄せられましたが、レビュー担当者が高く評価した出力コンテンツもありました。Meta (Llama) もそれほど良い結果は出せず、おそらく英語以外のコンテンツを最適化したり生成したりして信頼性の高い成果を得る段階に至っていないように思われます。出力品質は、今回の 1 か月の短期プロジェクトの間に向上が見られました。Bard では調査出力が改善され、ChatGPT ではドイツ語のコンテンツ制作が改善されました。ドイツ語の指示による ChatGPT の出力は、当初は 47/100 というスコアでした。

このスコアが、わずか 1 か月で 76/100 に上昇しました。英語の指示で生成された出力は、当初は 71/100 でしたが 1 か月で 86/100 に上昇しました (英語の平均スコアは 78.5/100)。

Bard も ChatGPT も、1 ~ 2 文ほどの短いコンテンツの生成に優れていました。短い文章に対しては否定的なフィードバックは非常に少ない結果となりました。

ここに示すスコアは、実験の 2 つのフェーズにおける長い形式のコンテンツに対するものです。

フェーズ 1 のスコア

ツール	使用できない	編集すれば使用可能	編集なしで使用可能	平均スコア
ChatGPT (ドイツ語による入力)	9	9	0	47.61
Bard (ドイツ語による入力)	2	14	2	68.33
ChatGPT (英語による入力)	3	9	6	71.67
Bard (英語による入力)	0	9	9	82.22
Meta (Llama) (英語による入力)	7	8	3	46.50
合計	21	49	20	

フェーズ 2 のスコア

ツール	使用できない	編集すれば使用可能	編集なしで使用可能	平均スコア
ChatGPT (ドイツ語による入力)	1	7	0	76.25
Bard (ドイツ語による入力)	2	4	2	66.25
ChatGPT (英語による入力)	0	4	4	86.875
Bard (英語による入力)	3	4	1	59.375
合計	5	19	7	

注: Meta (Llama) は、英語以外のコンテンツを最適化したり生成したりできる段階に至っていないため、フェーズ 2 では対象から除外しました。

詳しくはこちら

LIONBRIDGE.COM