

NoLangの機能分析と内製化開発の検討

NoLangの主要機能と提供仕様

NoLang（ノーラング）は、日本発のAI動画生成サービスであり、テキストや資料から自動でナレーション付きスライド動画を生成します¹。2024年7月の公開以降急成長し、2025年末時点でユーザー15万人・法人導入60社以上に利用されています²。以下、NoLangが備える主な機能・仕様を調査項目ごとに整理します。

1. 音声合成の精度・話者数・カスタマイズ性

- **多数の高品質AI音声:** NoLangには**300種類以上のAI音声**が搭載されており、さまざまな性別・年齢・声質・トーンのナレーションから選択できます³⁴。ビジネスシーンに応じて「重厚な低音」から「親しみやすい声」まで最適な声を使い分け可能です⁵⁶。日本語ナレーションの自然さについては、実際の利用者から「不自然な箇所は見当たらない」との評価も得ています⁷。
- **音声合成エンジンと自然さ:** 音声は最新の音声合成モデルにより**囁き声や抑揚**も表現可能で、感情豊かな読み上げが実現されています⁸。NoLangでは内部で複数の音声エンジンを統合している可能性が高く、ユーザーは「エンジン・性別・年齢・声質・用途・トーン」で検索し希望の声を絞り込めます⁴。例えばNoLang生成動画のサンプルでは**VOICEVOX**（オープンソースの日本語音声合成）による音声が使われており⁷、日本語話者にとって自然で聞き取りやすいナレーションとなっています。
- **話者のカスタマイズ（クローンボイス）:** **わずか1分程度の音声ファイル**からユーザー自身の声のクローンを生成できる機能があります⁹。日本語の録音からは日本語の、英語の録音からは英語のクローンボイスを作成でき、一度登録すればテキストを入力するだけで**自分の声で喋るナレーション**を何度でも生成可能です⁹。複数言語への対応も進んでおり、2025年末時点で**18言語対応**（英語・中国語・スペイン語など）を実現しています¹⁰¹¹。この多言語展開ではネイティブレベルの発音で音声生成されるとされています¹¹。

2. 字幕連動機能（自動生成・編集・エクスポート可否）

- **字幕の自動生成:** NoLangは音声ナレーションに対応した字幕（テロップ）を自動生成・配置します。特に新機能として、既存の音声付き動画ファイルをアップロードするだけで**音声認識&翻訳による多言語字幕付き動画**を生成できます¹²。例えば日本語の動画から英語字幕動画、英語音声から日本語字幕動画へと自動変換が可能で、18言語に対応しています¹²¹³。字幕の生成はAIが**音声認識→翻訳→字幕配置**まで一括で行い、ユーザーはファイルをアップロードして言語を選ぶだけで完了します¹⁴¹⁵。字幕付き動画への多言語変換もワンボタンででき、動画内容やタイミングを保ったまま別言語のナレーションと字幕に差し替えられます¹⁶。
- **字幕テキストの編集:** 自動生成された字幕テキスト（シナリオ）は、動画編集画面で**手動編集・調整が可能です**¹⁷。各シーンごとに字幕に表示するテキスト（ナレーション台本）を直接書き換えたり、音声読み上げの**漢字の読み方や固有名詞の発音修正**も行えます¹⁷。これにより専門用語の発音や表記揺れにも対応でき、ナレーションと字幕の精度を高めることができます。
- **字幕スタイルとエクスポート:** 字幕（テロップ）のフォント・サイズ・色や表示スタイルもカスタマイズ可能で、動画デザインに合わせて調整できます¹⁸。編集画面で字幕位置や表示タイミングの細かな調整もできるため、視認性の高い字幕付けが可能です。なお**字幕データの単独エクスポート（例：SRTファイル）**に関する直接の記述は確認できませんでしたが、完成動画をダウンロードすれば字幕は焼き付いた状態でMP4出力されます¹⁹。多言語字幕生成機能では、元動画にまず元言語字幕を付与し、その後ワンクリックで別言語字幕版の動画を生成するフローが示されており¹⁴²⁰、必要に応じ字幕付き動画を各言語分出力して利用できます。

3. アバター生成・挿入機能（動作の自然さ・カスタム可否・配置調整）

- ・**写真から生成するAIアバター**: NoLangの大きな特徴として、**ユーザーの写真1枚から本人そっくりのアバター**を作成する「リアルアバター」機能があります²¹。アップロードした顔写真をもとに、動画内で**本人が喋っているかのように動くアバター**を自動生成できます²¹。生成されたアバターは実写に近い口の動き（リップシンク）やまばたきまで再現され、あたかも本人がプレゼンしているような臨場感を演出します²¹。動作の自然さについては、口パクのタイミングが音声と合致し、瞬き・表情変化も加わることで**ロボット的な不自然さを抑えたリアルな挙動**を実現しています。下図はNoLangで生成されたアバターがスライド解説を行う動画の一場面、人物の顔が実際に口を開き語りかけている様子が確認できます。



図：NoLangで作成されたAIアバターがスライド資料の隅に登場し、ナレーションを行っている例（画像はプレスリリースより⁸²²）

- ・**アバターの種類とカスタマイズ**: NoLangでは2025年7月のアップデートで**100体以上のアバター**が新たに追加されました⁸。ビジネス向けのスーツ姿の人物からお待たせしてしまい申し訳ありません。調査に時間がかかりましたが、NoLangの機能詳細と自社開発の可能性について、以下の通りまとめました。

NoLangの機能分析と内製化開発の検討

NoLangの主要機能と提供仕様

NoLang（ノーラング）は、日本発のAI動画生成サービスであり、テキストや資料から自動でナレーション付きスライド動画を生成します¹。2024年7月の公開以降急成長し、2025年末時点でユーザー15万人・法人導入60社以上に利用されています²。以下、NoLangが備える主な機能・仕様を調査項目ごとに整理します。

1. 音声合成の精度・話者数・カスタマイズ性

- ・**多数の高品質AI音声**: NoLangには**300種類以上のAI音声**が搭載されており、さまざまな性別・年齢・声質・トーンのナレーションから選択できます³⁴。ビジネスシーンに応じて「重厚な低音」が

ら「親しみやすい声」まで最適な声を使い分け可能です⁵⁶。日本語ナレーションの自然さについては、実際の利用者から「不自然な箇所は見当たらない」との評価も得ています⁷。

- ・**音声合成エンジンと自然さ**: 音声は最新の音声合成モデルにより**囁き声や抑揚**も表現可能で、感情豊かな読み上げが実現されています⁸。NoLangでは内部で複数の音声エンジンを統合している可能性が高く、ユーザーは「エンジン・性別・年齢・声質・用途・トーン」で検索し希望の声を絞り込めます⁴。例えばNoLang生成動画のサンプルでは**VOICEVOX**（オープンソースの日本語音声合成）による音声が使われており⁷、日本語話者にとって自然で聞き取りやすいナレーションとなっています。
- ・**話者のカスタマイズ（クローンボイス）**: **わずか1分程度の音声ファイルから**ユーザー自身の声のクローンを生成できる機能があります⁹。日本語の録音からは日本語の、英語の録音からは英語のクローンボイスを作成でき、一度登録すればテキストを入力するだけで**自分の声で喋るナレーション**を何度でも生成可能です⁹。複数言語への対応も進んでおり、2025年末時点で**18言語対応**（英語・中国語・スペイン語等）を実現しています¹⁰¹¹。この多言語展開ではネイティブレベルの発音で音声生成されるとされています¹¹。

2. 字幕連動機能（自動生成・編集・エクスポート可否）

- ・**字幕の自動生成**: NoLangは音声ナレーションに対応した字幕（テロップ）を自動生成・配置します。特に新機能として、既存の音声付き動画ファイルをアップロードするだけで**音声認識&翻訳による多言語字幕付き動画**を生成できます¹²。例えば日本語の動画から英語字幕動画、英語音声から日本語字幕動画へと自動変換が可能で、18言語に対応しています¹²¹³。字幕の生成はAIが**音声認識→翻訳→字幕配置**まで一括で行い、ユーザーはファイルをアップロードして言語を選ぶだけで完了します¹⁴¹⁵。字幕付き動画への多言語変換もワンボタンででき、動画内容やタイミングを保ったまま別言語のナレーションと字幕に差し替えられます¹⁶。
- ・**字幕テキストの編集**: 自動生成された字幕テキスト（シナリオ）は、動画編集画面で**手動編集・調整が可能です**¹⁷。各シーンごとに字幕に表示するテキスト（ナレーション台本）を直接書き換えたり、音声読み上げの**漢字の読み方や固有名詞の発音修正**も行えます¹⁷。これにより専門用語の発音や表記ゆれにも対応でき、ナレーションと字幕の精度を高めることができます。
- ・**字幕スタイルとエクスポート**: 字幕（テロップ）のフォント・サイズ・色や表示スタイルもカスタマイズ可能で、動画デザインに合わせて調整できます¹⁸。編集画面で字幕位置や表示タイミングの細かな調整もできるため、視認性の高い字幕付けが可能です。なお**字幕データの単独エクスポート（例：SRTファイル）**に関する直接の記述は確認できませんでしたが、完成動画をダウンロードすれば字幕は焼き付いた状態でMP4出力されます¹⁹。多言語字幕生成機能では、元動画にまず元言語字幕を付与し、その後ワンクリックで別言語字幕版の動画を生成するフローが示されており¹⁴²⁰、必要に応じ字幕付き動画を各言語分出力して利用できます。

3. アバター生成・挿入機能（動作の自然さ・カスタム可否・配置調整）

- ・**写真から生成するAIアバター**: NoLangの大きな特徴として、**ユーザーの写真1枚から本人そっくりのアバター**を作成する「リアルアバター」機能があります²¹。アップロードした顔写真をもとに、動画内で**本人が喋っているかのように動くアバター**を自動生成できます²¹。生成されたアバターは実写に近い**口の動き（リップシンク）**や**まばたき**まで再現され、あたかも本人がプレゼンしているような臨場感を演出します²¹。動作の自然さについては、口パクのタイミングが音声と合致し、瞬き・表情変化も加わることで**ロボットのな不自然さを抑えたリアルな挙動**を実現しています。
- ・**アバターの種類とカスタマイズ**: NoLangでは2025年7月のアップデートで**100体以上のアバター**が新たに追加されました⁸。ビジネス向けのスーツ姿の人物からアニメ風キャラクターまで**Live2D形式の2Dアバター**や、初めて**リアルな3Dアバター**も搭載されています⁸。さらにユーザーが自分で用意した**PSD/Live2Dデータからオリジナルキャラクターをアップロード**してアバターにすることも可能になりました（このアップロード用クレジットも無料化）⁸。つまり、NoLang上では**自分の顔写真によるリアル分身**だけでなく、企業のマスコット等のオリジナルキャラを動かすことも容易になっています。加えて、アバターはシーン内容に応じて**笑顔・怒り・涙**などの表情変化を自動で付けること

もできます。こうした非言語コミュニケーションの表現力強化により、より感情豊かで人間らしい動画演出が可能です。

- **配置と複数アバター:** アバターの画面上での**位置やサイズ**は編集時に調整できます¹⁸。スライド画面の隅に小さく配置してプレゼンター風にしたり、画面いっぱいに登場させることも自由です。また、NoLangには**対話形式モード**があり、異なる2体のアバターが会話する動画も生成できます。実際の例では、2人のAIアバターが質問役と解説役に分かれて掛け合い形式で説明することで、内容がより明快になる効果が見られました。このように**複数アバターを同時に登場**させることも可能で、よりリッチな動画シナリオに対応しています。

4. 編集機能（シナリオ編集・タイムライン編集・プレビュー等）

- **生成後の編集画面:** NoLangは動画を自動生成するだけでなく、**生成後にユーザー自身で細かな編集を行える**点も優れています¹⁷。編集画面では以下のような操作が可能です¹⁷¹⁸：
 - 字幕テキスト（各シーンのナレーション台本）の直接編集
 - 漢字の読み仮名指定や難読語の発音修正
 - 各シーンで使用されている**画像や背景**の差し替え（アップロード画像への変更、あるいはAI画像生成での置換）¹⁸
 - BGM（バックグラウンド音楽）の変更、ナレーション音声の種類（声質）やスピード調整
 - アバターの差し替えや、アバター画像のサイズ・レイアウト調整¹⁸
- **タイムライン編集:** NoLang上で**動画の一部をカット（不要シーンの削除）**したり、シーン順序を入れ替えることもできます。生成直後に不要なイントロを切り落としたり、長過ぎる説明を間引くといった**簡易的な動画編集操作**ができるため、外部の動画編集ソフトに持ち出す手間を減らせます。また、動画長さ自体は生成前に30秒～3分の範囲で指定可能ですが、生成後に各シーンの表示秒数を伸縮させることで微調整も可能です（具体的なUI記載はありませんが、テキスト量に応じ自動調整もされています）。
- **プレビューと再生:** 動画生成は1分程度の動画なら**3～5秒ほど**で完了し、生成後すぐにブラウザ上で再生プレビューできます。編集画面でテキストや素材を修正した場合も、編集内容を即座に反映したプレビュー再生が可能です。必要に応じて再度「生成」処理を実行し、修正済みの最終動画を数秒～数十秒でレンダリングできます。つまり**動画の確認→編集→再生**のサイクルがNoLang内で完結し、リアルタイムに近い形で編集結果を確認できるわけです。
- **テンプレート機能と自動演出:** NoLang 4.0では**テンプレート機能**も追加され、ユーザーはデザインや構成のテンプレを選んで手軽に見栄え良い動画を作れるようになりました。また、シーンに合った**効果音の自動挿入機能**も実装され、例えば強調したい場面でボンと効果音が鳴るなど、細かな演出もAIが自動で行います。これにより編集作業なしでも飽きさせない動画効果を盛り込めます。将来的にはアバターのジェスチャー強化など、非言語表現の編集機能も拡充予定とされています。

5. 入力対応フォーマット（テキスト・画像・PDF・PowerPoint等）

- **テキスト入力:** 最も基本的な使い方はテキストからの動画生成です。NoLangでは、ユーザーが**動画にしたい内容の文章**をそのまま入力することで、それを台本として動画を生成できます¹。たとえば「○○の使い方を解説する動画」と一行入力すれば、AIが適切に構成を考えナレーション原稿や視覚素材を作ってくれます。専門知識不要で思いついたアイデアを即動画化できる点が初心者にも魅力です。
- **PDF・PowerPoint資料:** NoLangは**PDFやPPTX形式のスライド資料**を入力として受け取り、自動でその内容を解析して動画化することができます¹¹。プレゼン資料内のテキストや画像を抽出し、AIが要約・再構成してナレーションや字幕を生成します。実質、「既存の企画書や提案書をアップロードするだけで解説動画ができる」機能で、ビジネス活用で重宝されています¹¹。Chrome拡張機能経由ではWebページ上の文章も取得できるため、**Webサイトの内容を要約した動画**もワンクリックで生成可能です。英語ページでも日本語要約動画に変換できるので、リサーチ資料の理解などにも役立ちます。ただしこれら外部コンテンツ要約動画は権利上商用利用不可なので注意が必要です。
- **画像入力・活用:** 画像ファイルそのものを単独で動画化するモードはありませんが、NoLangでは**動画内で使用する画像素材**をユーザーがアップロードして差し込むことが可能です¹⁸。自社商品の写真

や特定のグラフ画像などを各シーンに当てはめ、よりオリジナリティある動画にできます。また、画像を自動生成する機能もあり、テキスト内容にマッチするイラストや写真をAIが生成・挿入してくれます。例えば「エナジードリンクの雑学」動画ではAI生成画像が使われましたが、内容にしっかり合致したビジュアルが得られていました⁷。このように**画像素材は自動選定/生成と手動アップロードの両方に対応**しています。

- **音声ファイル**: NoLang 4.0からの新機能として、**音声のみのファイル（例: ラジオ音源や録音データ）**をアップロードして動画化することが可能になりました。具体的には、ユーザーが手元の音声ファイルをNoLangに投入すると、その音声をナレーションとして利用しつつ、内容に沿った画像や字幕、場合によっては口パクアバターまで自動で付与した動画が生成されます。ポッドキャストや社内研修の録音データを映像付きコンテンツに再活用するケースなどが考えられます。
- **動画ファイル**: 同様に、既存の**動画ファイル（音声付き）**をアップロードして別形式の動画を生成することも可能です。アップした動画に対し、NoLangはまず元言語の字幕を自動生成します（文字起こし）¹⁵。さらにユーザーが希望すればボタン一つで他言語字幕版の動画に変換できます²⁰。この機能を使えば、社内用に撮影した日本語動画を即座に英語字幕付きにして海外支社に共有、といったことができます。対応フォーマットは、動画なら**.mp4 / .mov / .webm**、音声なら**.mp3 / .wav / .m4a / .aac / .ogg / .flac**に対応しています。
- 補足として、NoLangは**Chrome拡張**によるブラウザ連携も強みです。例えば閲覧中のWebページをボタン一つで日本語要約動画にするなど、入力元の多様さも魅力です。

6. 出力形式（MP4動画・埋め込みコード等）

- **MP4動画ダウンロード**: NoLangで生成した動画は**MP4形式**でダウンロードできます¹⁹。動画視聴ページの「ダウンロード」ボタンから即MP4ファイルを保存可能で、これを自社サイトやSNS、研修システムにアップロードして利用できます¹⁹。無料プランの場合、ダウンロードされるMP4にはNoLangのロゴ透かし（ウォーターマーク）が入りますが、有料Standardプラン以上では透かし無しでエクスポートできます。また、**横型16:9**だけでなく**縦型9:16**のショート動画形式でも出力でき、Premiumプランでは横→縦動画のワンクリック変換機能も提供されています。これは生成後にTikTok/YouTubeショート用のクロップを自動で行うもので、短尺PR動画の量産に便利です。
- **共有リンクと埋め込み**: NoLangの動画はクラウド上に保存され、**共有用のリンク**を発行することができます。動画画面の「共有」ボタンからアクセス範囲を「リンクを知っている人に公開」に設定すると、固有のURLが生成されます。このリンクを相手に伝えれば、NoLangのホスティング上で動画を視聴できます。埋め込みコードについて公式記述はありませんが、共有リンク先ページは一般的なWeb動画プレイヤーになっているため、必要に応じてiframe等で自社サイトに埋め込むこともできるでしょう（ただし閲覧にはNoLang側のサーバーにアクセスする形になります）。一部の競合サービス（Synthesiaなど）は独自の埋め込みプレイヤーを提供し、視聴者側で言語字幕を切り替えられる機能もあります。NoLangも将来的にそうしたマルチリンガルプレイヤーを用意する可能性があります。現時点ではシンプルな共有リンク提供がメインです。
- **その他出力**: 法人向けには、動画をeラーニング用のSCORMパッケージで出力する機能や、制作した動画の分析レポートを出す機能などが望まれますが、NoLangにそのような機能があるかは未公開です（SynthesiaはSCORMエクスポートや視聴分析ダッシュボードを持っています）。NoLang Businessプランでは専用のチーム共有機能があり、チーム内で動画を共有・編集できるよう。出力形式自体は基本MP4ですが、用途に応じて最適化された解像度・ビットレートでダウンロードされるものと思われます（SNS向け、プレゼン向け等、プリセットがある可能性があります）。

7. API連携（外部アプリ操作可否とドキュメントの有無）

- **NoLang APIの提供状況**: NoLangは**外部アプリから操作可能なAPI**を提供しています。2025年7月、NoLang 4.0のリリースと同時に「**NoLang API**」を**一般公開**したと発表されました。これにより開発者は自社システムやサービスからプログラム経由でNoLangの動画生成機能呼び出せます。公式FAQにも「API連携は可能ですか？ 有料プランでAPI利用が可能です。詳しくはAPIドキュメントをご覧ください」との記載があり、Businessプラン以上でAPI利用権が付与されるようです。

- **APIの機能範囲:** NoLang APIでは**テキスト・資料（PDF/PPT）・音声・動画**といったあらゆる入力から動画コンテンツを生成できます。業界別の活用例として、製造業なら「マニュアルPDFをアップロードして多言語対応の研修動画を作成」、金融なら「IR資料から経営陣の3Dアバターを載せたPR動画を量産」といったソリューションが紹介されています。実際にAPI経由で可能な操作は、NoLangのUI上で出来ること（アバター作成、音声クローン登録、動画生成、言語変換など）のほとんどをカバーしていると推察されます。プレスリリースにも「外部サービスや自社SaaSへ動画生成機能を組み込める」と明記され、たとえば**自社のIRサイトにAPI連携し、決算説明資料をアップすると即座に動画化してサイト上に公開**するといった自動化も可能とのことです。
- **ドキュメントと利用条件:** NoLang APIの詳細な技術ドキュメントは一般には公開されていないようですが、リリース告知内に**Docsへのリンク**が含まれていました。また開発者ハブ的な案内もあるようなので、法人契約者にはREST APIのエンドポイントやパラメータ仕様が提供されるものと思われます。利用には**APIキー/認証**が必要で、呼び出し回数や生成分数に応じてクレジットを消費する仕組みでしょう。ビジネスプランでは高度なセキュリティやチーム管理機能も含まれるとされ、API利用によるデータ連携も安全に行えるよう配慮されていると考えられます。なお今後の展望として、NoLangはCRM（顧客管理）やMA（マーケティング自動化）ツールとのAPI統合を更に進め、顧客ごとに内容を変えた動画を自動配信する、といった**動画生成のインフラ化**を目指すと表明しています。

8. 商用ライセンス利用可否（商用利用制限・ライセンス体系）

- **商用利用の可否:** NoLangで作成した動画は、基本的に**商用利用可能**です。ただしコンテンツによって適用ライセンスが異なります。NoLangが生成する通常のAI動画（プロンプト入力や資料要約で生成したもの）は**クリエイティブ・コモンズCC BY-SA（表示・継承）ライセンス**の下で提供されます。つまり「**動画生成：NoLang**」等の**クレジット表示**を行い、かつ同じ**CC BY-SA条件で公開する限り改変・営利利用が自由**という形です。このためYouTube投稿、広告利用、企業SNS掲載なども問題なく行えます。実際、NoLang利用ガイドでも「条件を満たせば作成動画を商用利用可能」と明記されています。重要なのはクレジット表記で、動画の最後や説明欄に「Video generated by NoLang (Mavericks)」等の記述を入れる必要があります。
- **注意点（要約動画の権利）:** ただし**Chrome拡張でWebページを要約して生成した動画**については注意が必要です。この場合、元のWeb記事の著作権をユーザーが持っていない限り、**生成動画を商用利用したり他の媒体へ転載・投稿することはできない**と案内されています。要するに、他人の著作物を無断で要約し公開すると二次的利用の問題があるため、NoLangとしても商用利用NGとしています。従って社内研修など**クローズドな用途なら問題にならない**でしょうが、一般公開・商用利用する場合は自社所有コンテンツから生成するのが安全です。
- **料金プランとライセンス体系:** NoLangは**基本利用無料**で開始でき、有料プラン（月額制）で拡張機能や生成枠が増える**サブスクリプション型ライセンス**です。主要プランは次の通りです：
 - **Freeプラン:** 月額0円、毎月200クレジット付与（約20本の動画生成相当）。アバター付き動画は2本/月まで等の上限あり。生成動画には透かし（ウォーターマーク）付き。
 - **Standardプラン:** 月額2,980円、2,000クレジット（動画200本程度）。アバター動画は20本/月まで、AI画像生成も10本分まで利用可。透かしなしダウンロード可、縦型変換機能あり。より高度な編集や追加機能が使える。
 - **Premiumプラン:** 月額7,980円、7,000クレジット（動画700本程度）。アバター動画70本、AI画像35本まで。Standardの全機能に加え大量生成向け。
 - **Businessプラン:** 月額は**要問い合わせ**（カスタム）。クレジット・生成数は応相談で増量可能。企業向け高度機能（専用サポート、チーム管理ツール、API利用、独自機能相談など）が提供される。
- **ライセンス条件:** 前述の通り、生成コンテンツにはCC BY-SAが適用されますが、**Businessプランなどでは別途契約で緩和可能な可能性**もあります。例えば自社社員の顔写真から作ったアバター動画であれば、自社コンテンツとして第三者へのCC共有を避けたい場合もあるでしょう。その際はMavericks社と個別契約でクレジット表記を省略したり、専用ライセンスで提供してもらう交渉も考えられます（Synthesiaなど競合ではEnterprise契約で独自ライセンス提供するケースがあります）。NoLang公式サイトでは**商用利用時はクレジット表記を忘れない**ようにとの注意だけなので、通常はCC表示で十分と考えられます。

- ・ **オンプレミス提供の有無**: 現在のところ、NoLangはクラウドサービスとして提供されており、ソフトウェアを購入して自社サーバーに導入する形のライセンス提供は行っていないようです。法人向けでも「NoLang for Business」としてクラウド上の専用環境やサポートを提供する形で、**あくまでSaaSとしての利用契約**になります。したがって、社内システムに組み込む場合でもAPI連携によるクラウド利用となり、自社完結のライセンスはありません。ただし、動画生成AIを自社システムに組み込む受託開発などもMavericks社は相談に応じる旨を述べており、要件次第ではカスタムライセンス契約も可能かもしれません。

以上がNoLangの機能・ライセンス面の詳しい分析です。NoLangは**無料でも使えて商用利用OK（クレジット要）**という手軽さと、音声・字幕・アバターを組み合わせた高度な自動編集機能で、**日本企業の動画活用を革新しているサービス**と言えます。

同等システムを自社開発（内製）する場合の検証

NoLangのような音声付きスライド動画自動生成システムを、自社で内製開発・運用する場合の**技術的実現性**と**費用対効果**を検討します。ポイントは、NoLangが持つ各機能を実現するための要素技術を揃えられるか、またそれにかかるコストと時間が現実的かという点です。

内製開発に必要な主な技術要素

NoLang相当の機能を内製するには、大きく以下のコンポーネントが必要です：

1. **文章解析・シナリオ生成AI** – 入力されたテキストやPDFから動画用の構成案やナレーション台本を作るAI。NoLangではChatGPTのような大規模言語モデルで実現している部分です。自社でも**OpenAI GPT-4 API**などを使えば比較的簡単に再現できます。例えば、ユーザーが入力した講座テキストから要点を抽出して3～5シーン程度の脚本を生成する処理です。日本語対応も含め、OpenAIやAzure OpenAI Serviceなど既存APIに頼るのが現実的で、自前で言語モデルを訓練する必要はありません。
2. **テキスト音声合成（TTS）** – 台本テキストをナレーション音声に変換するエンジン。高品質なものを目指すなら**クラウドのTTSサービス**利用が手堅いです。例えば**Google Cloud Text-to-Speech**や**Amazon Polly**、**Microsoft Azure Speech**は多言語・多彩な声質を提供しており、API経由で簡単に音声ファイルを取得できます。日本語女性・男性はもちろん、英語など主要言語も網羅しています。NoLangも内部ではVoicevox等のOSS音声とクラウド音声を組み合わせて300種用意したと推測されます³⁷。自社開発なら、**VOICEVOX**（無料の高品質日本語音声）を組み込むのも良い選択です。VOICEVOXはオープンソースで商用利用可、標準で8種ほどの日本語声があり、追加ボイスも豊富です。実際NoLangの日本語音声はVoicevox由来の声で自然な仕上がりでした⁷。一方、多言語音声はGoogleやMicrosoftのTTSで補えば、主要言語はカバーできます。**コスト**もTTSは安価で、例えばGoogleでは日本語1文字＝約0.02円程度（100万文字16ドル）なので、5分のナレーション（750文字程度）でも15円程度です。
3. **クローンボイス（話者の声合成）** – ユーザーの肉声でナレーションする機能。技術的には**Voice Conversion**や**Few-Shot音声合成**が必要で、これは難度が高めです。OSSでは「VITS+ContentVec」を用いたクローン音声生成などがありますが、品質的にNoLangレベルにするにはハードルがあります。代替として**Resemble AI**や**Microsoft Custom Neural Voice**など商用APIがあり、数十秒～数分のサンプル音声で話者声モデルを作れます。ただ費用が高かったり手続きが必要です。**内製初期段階ではクローンボイス機能は省き、後述する外部サービスSynthesia等のAPI活用で補う**のも手です。Synthesia Enterpriseではユーザーの顔だけでなく声もクローン化できますが、高額なため簡易には導入しづらいかもしれません。
4. **アバター生成・アニメーション** – これが内製でもっとも難しい部分です。NoLangの強みである「写真1枚からリアルに喋るアバター」をゼロから再現するのは極めて高度な技術開発になります。考えられるアプローチは：

5. **外部API利用:** 例えば**D-ID社のAPI**は、入力画像と音声から高精度な話者動画を生成できます。D-IDは既に150万以上の動画生成実績があり、レンダリング速度も実時間の4倍（100FPS）と高速です。プランにもよりますが月額数百～数万円で利用でき、内製で苦労するより合理的です。**Synthesia API**も検討対象ですが、Synthesiaは自社でWebアプリを完結させるビジネスモデルのため、APIは主に大企業向けで費用も高いです（参考：Synthesia Starterプランは月10分まで動画生成\$29²³）。中小規模であれば、比較的安価な**HeyGen**（APIあり、175言語対応）や**Colossyan**（PPTアップ対応で教育向き）も選択肢です。これらサービスはいずれも**AIアバターの生成をオンデマンドで提供する**もので、自社システムからAPI呼び出し→生成動画取得の流れで組み込みます。例えば**D-IDではAPIキー発行後、POSTリクエストで画像とテキスト/音声を送ると動画URLが返ってくる**仕組みで、技術文書も整っています。
6. **オープンソース技術組み合わせ:** Wav2Lip（口パク合成）＋First Order Motion Model（汎用画像アニメーション）といったOSSを組み合わせ、静止画から動画を合成する手もあります。Wav2Lipは音声波形に合わせて静止画の口を動かすモデルで、精度は高いですが口周りのみの変化です。FOMMを使えば首振りや簡単な表情変化も可能ですが、設定が難しく品質も保証できません。さらに瞬きの自然さなど細部を詰めるには追加処理が必要です。結果的に相当のディープラーニング知見と開発リソースを要します。**最新の研究モデル**（e.g. PIRenderer, SadTalker等）を追って導入することも可能ですが、安定稼働させるには手間がかかるでしょう。
7. **2Dキャラクターや3Dモデルの活用:** 自社開発の別のアプローチとして、アニメ風キャラであればLive2Dモデルを動かす、中の人に3Dモーションキャプチャしてもらう等が考えられます。しかしこれらはコストや手間が大きく、せっかく自動化したいのに結局手作業が必要になり本末転倒です。やはり**現実的にはD-ID等既存AIサービスに任せるのが無難**です。Synthesia系サービスは「**驚くほど自然なAIアバター**」を売りにしており、細かな瞬き・首振りなどマイクロジェスチャーも備えています。NoLangも今後そうした微細動作を実装予定と述べています。これらを自社一社で追従するのは難易度が高いと認識すべきでしょう。
8. **複数アバター:** 内製でも、スピーカーを複数用意して会話させること自体は可能です。シナリオ上でセリフA,Bと話者を振り分け、それぞれ別のアバター動画を生成・合成すればよいので、技術的には単一アバターの延長です。ただし画面レイアウト設計（例えば左右に2人物を配置等）や、発話順序の管理が必要です。NoLangの対話モードでは2人が交互に話す脚本をAIが書き、それに沿って2つのアバター動画を合成していました。自社でも似たロジックを組めますが、まずは**単一アバター生成を安定させる**のが先決です。
9. **映像編集・合成処理** – 最終的に字幕テキスト、画像、アバター動画、音声、BGMなどを一つの動画にレンダリングする工程です。NoLangではクラウド上でAIが自動編集していましたが、内製では**FFmpeg**など動画処理ライブラリを活用して実装します。各シーンごとに背景画像の上にテロップを描画し、さらに隅にアバター動画（クロマキー合成か円形切抜きなど）をオーバーレイし、ナレーション音声とBGMをミックスする、といった処理になります。FFmpegはコマンド一発でもテキストや画像重畳、音声ミックスが可能なので、自動生成スクリプトを組めば対応できます。また、動画編集の操作をユーザーにさせる場合、ウェブ上で**タイムラインUI**を提供する必要があります。例えばシーン単位にプレビュー再生や、字幕テキストのタイミング調整などです。Web上で動画編集を可能にするOSSフレームワーク（例：MoviePy+Flaskでサーバーサイド処理、ブラウザは軽微な操作のみなど）の採用も検討できますが、UI部分はかなりの実装労力です。内製初期は**自動生成結果のみダウンロード（編集不可）**でも良いかもしれません。徐々に編集機能を拡充する方が段階的アプローチとして現実的です。

以上、必要技術を洗い出しました。総じて、**音声合成や字幕生成は既存技術で賄えますが、アバター部分は外部サービスの活用が鍵**となりそうです。また、全体を統合するアプリケーション部分（フロントエンドUIやワークフロー構築）も見逃せない工数です。NoLangのように洗練されたUXを作るには相応のWeb開発力が必要になります。

利用可能なOSSやクラウドAPIの具体例

上記技術の候補として使える主なOSS/クラウドサービスをまとめます：

- **LLM/要約:** OpenAI GPT-4 API（高精度な日本語要約が可能）、ChatGPT Plugins/Tools（Web要約用）、または自社でSudachiなど形態素解析＋ルールベース要約でも可。
- **音声合成:**
 - クラウドAPI: Google Cloud TTS、Amazon Polly、Microsoft Azure Speech（いずれも高品質多言語。Azureは約400種類の音声とスタイル表現対応）。
 - OSS: VOICEVOX（日本語8声+有志追加多数、リアル系からアニメ声まで。NoLang実績あり⁷）、OpenJTalk（日本語ベース、品質中程度）、Mozilla TTS/Coqui TTS（英語他いくつかの事前学習モデル有り、カスタム学習も可）。
- **音声認識&翻訳:** Whisper (OpenAIの高精度音声認識モデル、多言語対応)、Google Cloud Speech-to-Text、DeepL API or Google Translate API（字幕翻訳用）。NoLangは「翻訳×動画編集の自動化」を謳っており、これらの組み合わせで実現しています¹²。
- **アバター生成:**
 - クラウドAPI: D-ID API（1枚画像→人物動画。Trialあり低コスト）、Synthesia API（要エンタープライズ契約）、HeyGen API、Colossyan API、Soul Machines（高リアリティだが重厚）、Hour One（Wix買収でサービス形態変化）。
 - OSS: Wav2Lip（音声リップシンク）＋First Order Motion Model / SadTalker等（顔全体アニメーション）。Unity+Live2D SDK（2Dアバター表現に）、Three.js+VRM（3DモデルをWebで動かす）なども応用可能。
- **画像/映像素材:** Pexels API・Unsplash API（フリー写真検索）、Stable Diffusion系API（画像生成で独自イメージ挿入）、Lorem Picsum（ダミー画像）、Storyblocksやニコニ・コモンズ（BGM/SE素材）など。
- **編集・合成:** FFmpeg（動画レンダリングの中核）、Canvas API / WebGL（ブラウザレビュー用）、Node.jsのfluent-ffmpegライブラリ（サーバーサイドFFmpeg制御）、Python MoviePy、SHOTSTACK API（動画合成クラウドサービス）など。

注意: こうしたOSS/APIを組み合わせれば、かなりNoLangに近い機能群が揃います。しかし重要なのは、それらを一貫したワークフローに統合し、ユーザーが簡単に使えるようにすることです。OSSは設定やチューニングが必要なものも多く、クラウドAPIはそれぞれ認証や料金管理が別々に発生します。内製システムでは、その裏側をすべて隠蔽してシームレスに動かす工夫が求められます。

開発・運用コストの試算

初期開発コスト: 前述したように、内製開発にはマルチな技術スキルが必要で、開発期間も長めです。社内にAI開発チームがありリソース投入できるならよいですが、ゼロから外部ベンダーに委託すると**数千万円規模**のプロジェクトになる可能性があります。最低限、音声合成と簡易アバターだけの動画生成ツールでも、設計・実装・テスト含め**数百万円**は見ておくべきでしょう。特に**アバター部分は外部API頼み**と割り切れば自前開発を省けますが、それでもAPI連携の実装や試行錯誤は必要です。

ランニングコスト: 内製システムの維持費として、サーバー代やAPI利用料が発生します。ケーススタディとして、**月に10本・各5分の研修動画を生成**する場合を考えます： - テキスト要約（GPT-4）：1本あたり数円～数十円。10本で数百円。

- 音声合成（Azure音声など）：5分ナレーション～750文字→数円/本。10本で数十円。無視できるほど安い。
- アバター生成（D-ID）：プランによるが、高画質5分動画×10本で数千円～1万円程度か。Synthesiaなら10本=50分でStarter枠超過、Enterprise契約が必要（割高）。
- ストレージ/転送：動画10本（例えば1080p5分で各50MB）で計500MB。クラウド保存＋配信しても数円～数十円。社内利用で帯域限定なら negligible。
- サーバー運用：自前でレンダリングサーバーを動かすなら、GPUクラウド（月額5～10万円）を借りるか、

ローカル高性能PCを用意（初期数十万円）する必要も。もっともFFmpeg合成程度ならGPU不要で、CPUサーバー（月1万円以下）でも済みます。

以上概算では、**変動費は1本あたり数百円以下**に抑えられます。アバターをどのサービスに委ねるかが費用のボトルネックです。Synthesiaのような高品質サービスは便利ですが**価格が高く**、内製システムで多用するとNoLangの有料プランより割高になる恐れがあります。D-IDやHeyGenは比較的安価なので現実的選択肢でしょう。いずれにせよ**動画生成数が少なければクラウドAPI払い、数千本以上量産するなら自前運用**といったコスト構造です。

内製 vs 外注の比較: - **内製のメリット:** カスタマイズ自由度（社内教材のスタイルに完全適合）、ランニングコストを自社コントロールできる、社内ノウハウの蓄積。デメリットは技術ハードル・初期費用・人材確保。

- **外注（または外部サービス利用）のメリット:** 速攻性（すぐ使える）、信頼性（実績あるサービス品質）、アップデート任せられる。デメリットはカスタム要件の非対応、長期費用累積、データを外部に預ける不安。

両者の併用も一つの解です。例えば**短期的にはNoLangサービスを使い、ノウハウを得てから段階的に内製に切り替える**戦略です。NoLangは無料枠も大きく商用OKなのでまずPoCとして活用し、社内で自動動画の有用性を確認できます。その上で、どうしても内製化したい部分（例えば自社社員の顔を使ったアバター動画を外部クラウドに置きたくない等）が出てきたら、そこを重点的に自社実装に切り替える計画を立てると良いでしょう。

英語対応拡張性

将来的に英語など多言語の講義動画を自動生成するニーズにも触れておきます。NoLangは既に全18言語に対応し、日本語資料から英語動画を即生成するデモを示しています¹¹。内製システムでも、**多言語対応は比較的实现容易**です：- 音声合成はGoogle/AzureなどクラウドTTSで各国言語のナレーションを得られます。品質もネイティブに近い発音です。Synthesiaは140言語・2000以上の音声を提供しており、多言語展開は技術的には確立されています。

- スクリプト翻訳はDeepLやAzure Translatorで高精度にできます。専門用語辞書を整備すれば誤訳も防げます。NoLangの多言語機能も、資料を翻訳→現地語音声・字幕に置換する流れです。

- アバターの口パク同期も、D-ID等は100以上の言語でリップシンク調整されています。Wav2Lipも言語非依存なので問題ありません。

- UIやフォント面の国際化に留意すれば、英語はもちろん、中国語やスペイン語など主要言語はすぐカバーできます。右から左に書くアラビア語などは字幕表示の実装を工夫する必要がありますが、特殊ケースです。

要するに、**多言語サポートは音声合成・翻訳APIを使うことでスムーズに追加可能**です。自社の自動動画ワークフローを整備する際も、初めから英語字幕機能などを組み込んでおけばグローバル展開に対応できます。NoLangはまさにそれを売りに海外商談の効率化を図っています¹¹。

結論・提言

NoLangは音声合成の自然さ、字幕・アバターの自動統合、編集UIの使いやすさなど、非常に完成度の高いサービスです。それと同等のシステムを自社開発することは**技術的には可能ですが、大きな投資と開発期間を要する**チャレンジです。特にアバター生成の部分は、NoLangも自社のコア技術としており、同レベルを独自に構築するのは簡単ではありません。

自社開発の実現性についてまとめると:

- 音声合成・字幕生成など**テキスト処理系は比較的容易**に実装可能。既存APIやOSS利用で品質も確保できる。
- アバター・映像合成など**視覚処理系はハードルが高い**。外部のAIサービス(API)を活用することで補完すべき。
- **費用面**では、内製初期費用は大きいですが、長期的利用や社内資産化を考えるとメリットもある。逆に短期・少量利用ならNoLangのようなサービスを使った方が安上がり。
- **内製 vs サービス利用**の折衷案として、まずNoLangや競合サービスを活用し、ニーズを把握した上で徐々に内製に切り替えるのがおすすめ。必要に応じAPI連携で自社システムに組み込むこともできる。
- 将来的な英語対応も見据え、使用する技術・サービスは多言語対応力のあるもの（Google/Azure TTSやD-ID APIなど）を選ぶと良い。

自社講座・研修動画の自動生成ワークフロー構築の目的に対しては、**まずは現行ソリューション（NoLangなど）の活用でスピーディに成果を出し、その効果を測定**することが現実的でしょう。その上で、内製化によってどの部分で**独自価値（コスト削減や機密保持、カスタム機能等）**を出せるかを見極め、段階的に取り組むことを提言いたします。例えば、「社内資料から自動動画生成」という流れ自体はNoLangで十分可能なので、しばらくはNoLangビジネスプランを利用し、社内フィードバックを集める。その中で「もっとこうしたい」が出てきたら、API連携や内製ツールで補完する、といったハイブリッド戦略がリスクも低くおすすめです。

最後に、本調査の要点をまとめると以下の通りです:

- **NoLangの性能:** 音声合成は高品質で話者数も豊富、字幕・アバター・編集が統合された完成度の高いサービス。商用利用も容易で多言語にも対応 ³ ¹²。
- **内製の可能性:** 技術的要素はそろえられるが、特にアバター部分は既存APIの活用が現実的。VoicevoxやD-ID等、OSS/クラウドを組み合わせれば主要機能は構築可能 ⁷。
- **費用とリソース:** 内製初期費用は大きいですが、ランニングコストはコントロール可能。外部サービスは初期0で始められるが、長期では料金増も。自社の動画利用量や重要度によって判断する。
- **戦略:** 内製と外注（既存サービス）のハイブリッドでメリットを享受する。まずはNoLang等でスモールスタートし、段階的内製で自社に最適化したシステムを目指す。将来の英語展開も視野に、選定技術は国際対応可能なものを選ぶ。

以上、NoLangの詳細機能とそれを踏まえた自社開発の検討結果となります。自社の目的（講座・研修動画の効率的な大量生成）に照らし、最適な導入・開発プランの策定に本調査結果がお役に立てば幸いです。

参考文献・出典:

- NoLang公式サイト・プレスリリース（多言語対応、新機能発表） ¹¹ ³ ⁹ ¹⁶
- 導入事例・解説記事（Aismiley, AI Walker 等） ⁷ ¹⁷
- Mavericks社 note 解説記事（音声300種追加、検索機能等） ⁵ ⁴
- NoLang 4.0 API公開プレスリリース ⁸
- NoLang字幕機能プレスリリース ¹² ¹⁴
- 外部AI動画サービス情報（Synthesia vs D-ID 比較記事等） ²³
- その他出典（NoLang FAQ、各種サービスサイト）

¹ ⁷ ¹⁸ ¹⁹ 動画生成AI「NoLang（ノーラング）」の使い方と料金を解説！

https://aismiley.co.jp/ai_news/how-to-use-nolang/

2 3 9 10 11 16 21 動画生成AI「NoLang」、全18言語への多言語対応を開始。日本語のプレゼン資料から「自身のAIアバター」が流暢に喋る海外顧客向けの営業動画を即座に生成 | 株式会社Mavericksのプレスリリース

<https://prtimes.jp/main/html/rd/p/000000067.000129953.html>

4 5 6 動画生成AI「NoLang」、300種類以上の「音声」を追加し、動画の表現力を大幅に拡張。あらゆるビジネスシーンにおける最適な音声が可能に | Mavericks, Inc

https://note.com/note_maverick/n/nedd457f01d7e

8 22 日本発の動画生成AI「NoLang 4.0」& NoLang APIがリリース — 国内初 テキスト、資料、音声、動画すべてに対応 | 株式会社Mavericksのプレスリリース

<https://prtimes.jp/main/html/rd/p/000000010.000129953.html>

12 13 14 15 20 動画生成AI「NoLang」、動画・音声ファイルをアップロードするだけで「多言語字幕付き動画」を自動生成する新機能を搭載。日本語の音声を英語字幕にするなど、AIが「翻訳×動画編集」を完全自動化 | 株式会社Mavericksのプレスリリース

<https://prtimes.jp/main/html/rd/p/000000068.000129953.html>

17 NoLang(ノーラング)とは？特徴や料金、使い方まで完全解説！ | AI Walker

<https://walker-s.co.jp/ai/what-is-nolang/>

23 5 Best D-ID Alternatives In 2026 (Detailed Comparison)

<https://www.synthesia.io/post/best-d-id-alternatives>