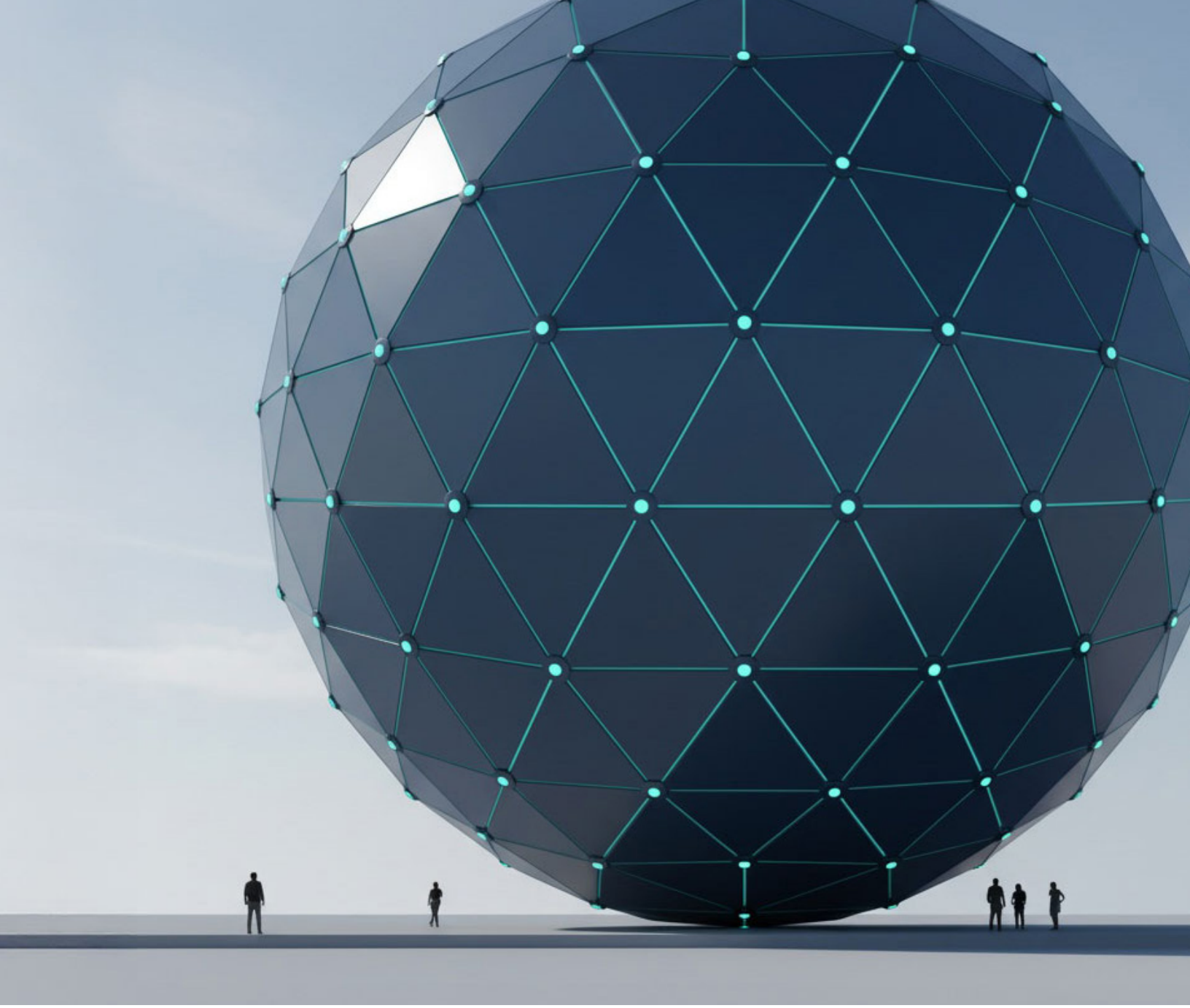


elastic

エアギャップ環境のための高度なベクトル検索

公共部門は、生成AI導入の重要な段階にあります。組織は、国全体の戦略や利害関係者の期待に後押しされ、より大規模なAI施策を展開する必要に迫られています。同時に、生成AIと検索は絡み合っています。ユーザーは、クエリに対する迅速な回答を求めて、リンクのリストを提供する検索エンジンの代わりに、対話型の生成AIを利用するようになっています。

公共部門でこれらの期待に応えるには、いくつかの課題が伴います。

1 生成AIを使用すること自体は可能ですが、十分な開発者がいません。組織は、最も効率的な結果を生み出すために戦略的にリソースを投資する必要があります。

2 公共部門の組織は、膨大な量のデータを扱っています。AIの基盤として使用できるよう、拡張性があり、高速かつ効率的なナレッジベースが必要です。

3 データは常にAIに対応しているわけではありません。重要な情報はPDF、図、ビデオなどの形式で格納されることが多く、そこから関連情報を抽出するのは難しい場合があります。

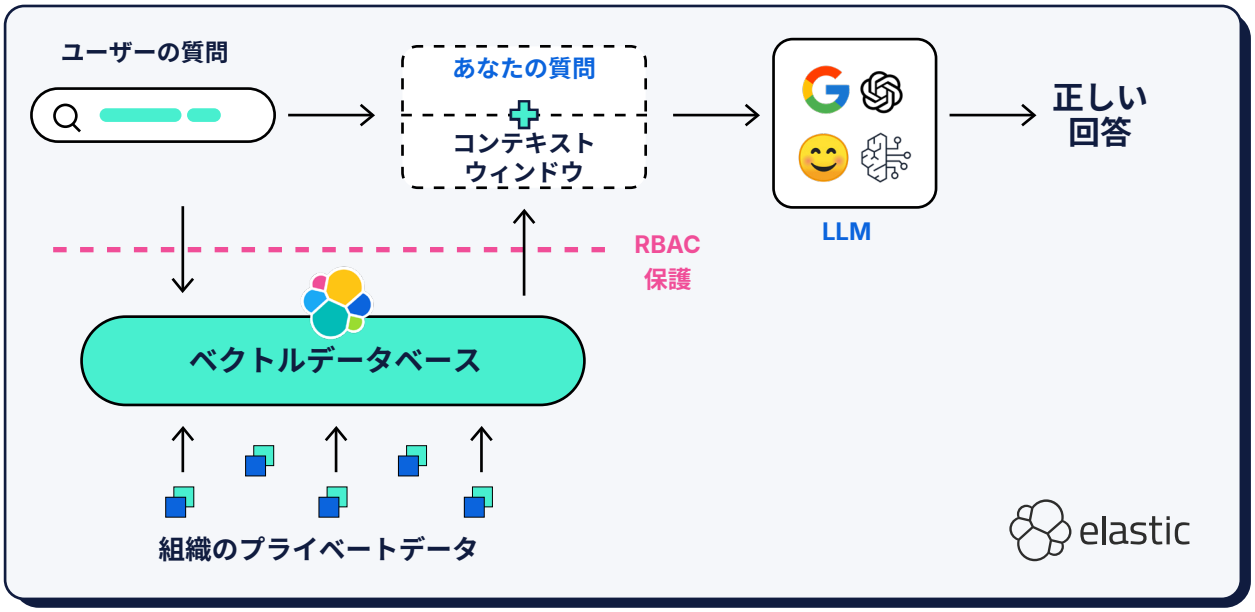
防衛および情報機関は、エアギャップ環境や技術キットで運営するという追加の課題に直面しています。

RAG：AIのための効率的なデータ検索モデル

多くの公共部門の組織は、関係者が行うクエリに合わせて大規模言語モデル（LLM）をトレーニングまたは微調整することが難しいと感じています。たとえデータが複雑であっても、独自のデータをAIですぐに使用したいと考えています。

 © 2025 Elasticsearch B.V. All rights reserved.

Retrieval-Augmented Generation（RAG）は、限られた時間とリソース、そして膨大な量の異種データをどのようにスケールするかという、より大きな課題に対処する拡張性のあるソリューションです。ElasticのRAGモデルでは、ユーザーの質問がLLMに送信される前に、まずベクトルデータベースにクエリが送られます。これにより、ベクトル検索を通じて組織の所有データから関連するコンテキストを抽出できます。このステップ、つまりコンテキスト層により、LLMがその出力に依存するデータの範囲が大幅に削減されます。エージェント型AIは、ユーザーとRAGの間のAIエージェントとクエリの相互作用を通じてこのモデルの性質をさらに大きく変えます。

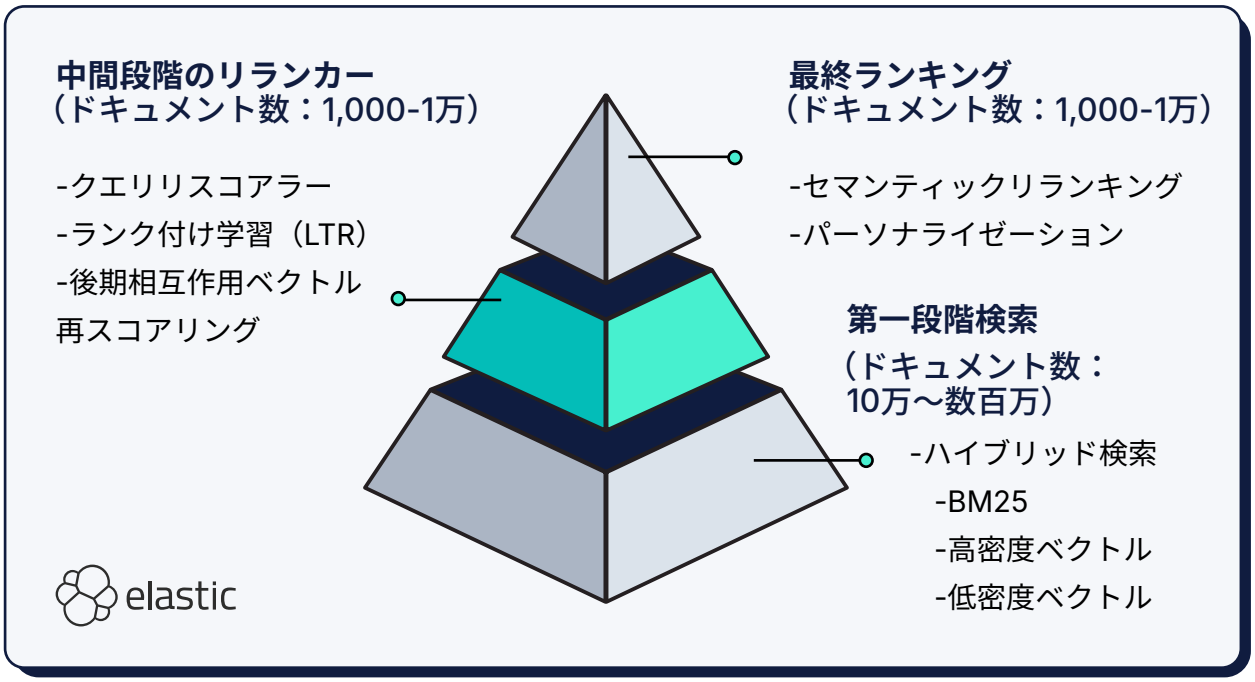


複雑なデータ型に対する多段階検索とマルチモーダル検索

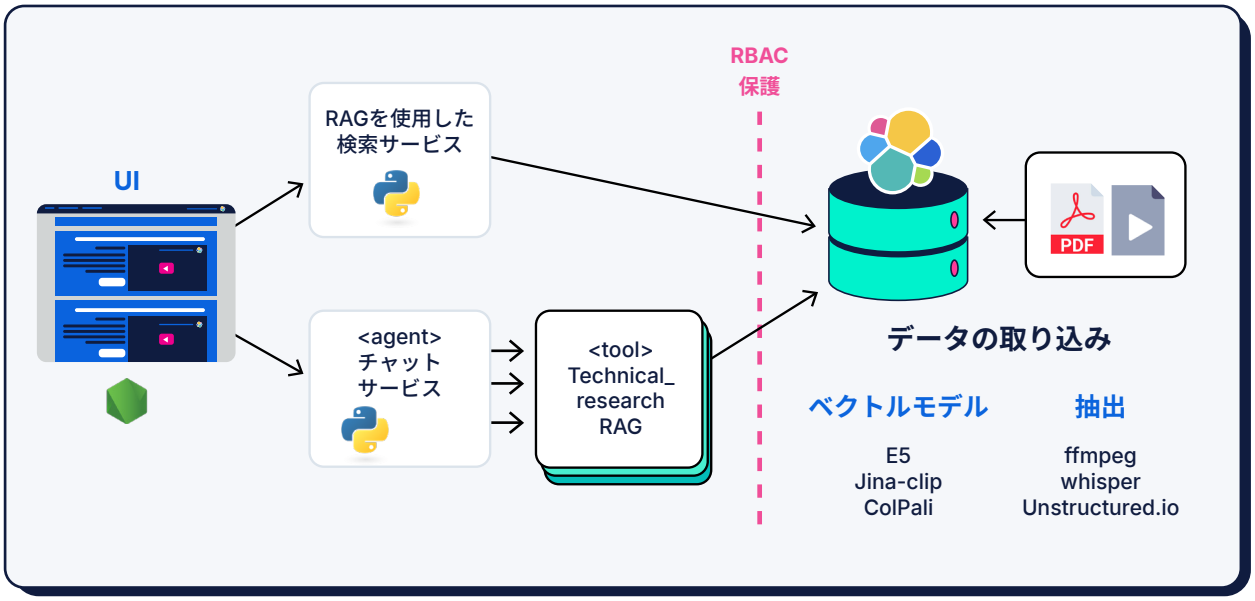
公共部門の組織がRAGアプローチで生成AIに複雑なデータを活用できるようにするため、Elasticは**多段階検索**と**マルチモーダル検索**を活用しています。

多段階検索は、RAGジャーニーにおける複雑なクエリを小さなタスクに分解し、最終的な回答を提供します。このアプローチは、おおよその結果と具体的な結果を得るためには異なるテクノロジーが必要であるという考えに基づいています。とりわけ、Elasticは、セマンティックリランキングのエアギャップオプションが完全にサポートされている最初のベクトルデータベースです。

 © 2025 Elasticsearch B.V. All rights reserved.



同様に、組織がさまざまなデータの情報を取得するためのより直感的で強力な方法を求める中、マルチモーダル検索のユースケースが急速に勢いを増しています。例えば、テーブルを含むPDFや、音声とビジュアルのコンポーネントを含むビデオなどです。マルチモーダル検索は、テキスト、画像、音声、ビデオのインプットを組み合わせて、最終的にAIアプリケーションが適切なコンテキストで質問に迅速に答えることを可能にします。

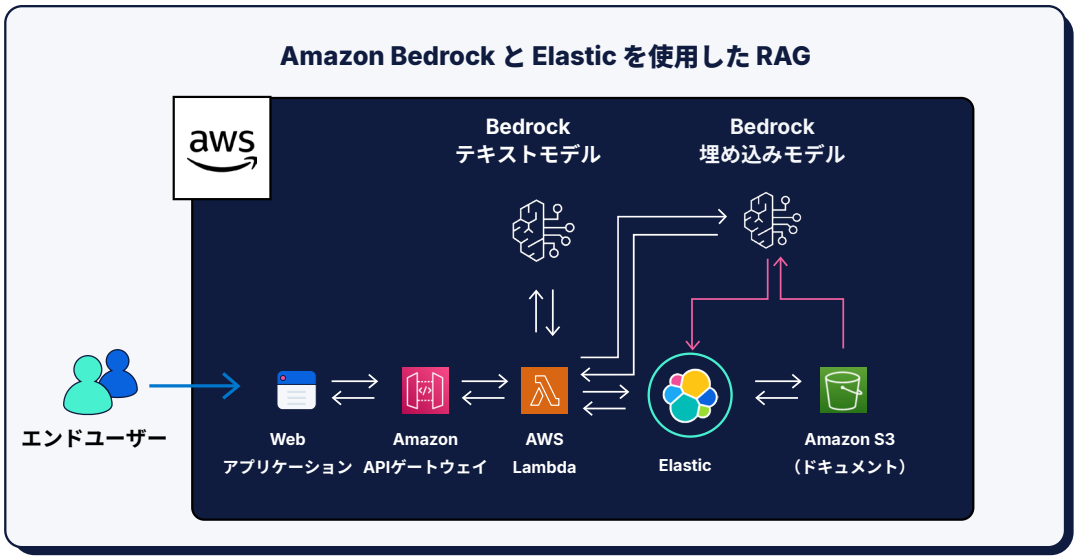


上記の構造を使用すると、組織は、120分の.MOVファイルの奥深くに埋め込まれた視覚情報や、1997年のPDFの63ページにある図のフレーズを見つけることができます。この機能により、複雑な独自データを総合的に検索でき、生成AIツールが出力に使用できるようになります。Elasticはサブイメージ検索も可能で、画像全体やビデオの

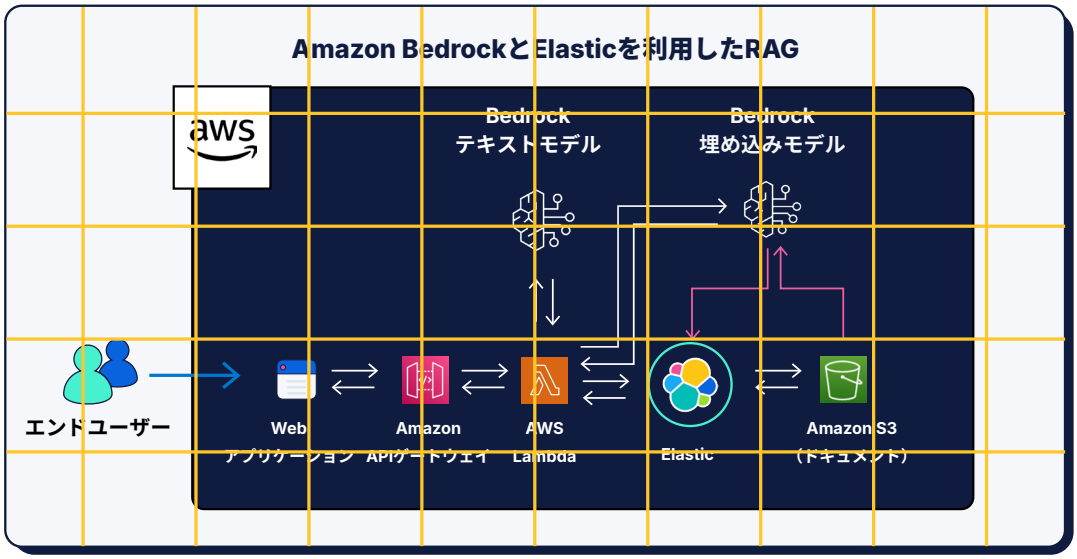
 © 2025 Elasticsearch B.V. All rights reserved.

フレームの比較的小さな部分を占めるものであっても、ユーザーのクエリとの全体的な関連性に基づいて検索結果で上位にランク付けされます。何かがオーディオトラックで議論されていて画面に表示されていない場合でも、またはその逆の場合でも、ユーザーの質問に関連していれば、Elasticはそれを見つけて検索結果で上位にランク付けできます。

RAGを使用した高度な画像検索



ColPali マルチベクトル



RAGアプローチによって画像ファイルから関連情報を表示する方法の例を具体的に見てみましょう。ここでは、「Amazon Bedrockを利用したElasticアーキテクチャの図」を検索したいとします。上記の最初のスライドから情報を取得するには、OCRとテキスト抽出を実行し、データがたどるパスを保持するためにフローを作成する必要があるでしょう。特に各文書が異なる場合、複雑になる可能性があります。また、8ページ以上に及ぶ

 © 2025 Elasticsearch B.V. All rights reserved.

複数の列と表を含むページ数の多い政府のPDFを検索する場合、要求される作業ははるかに大きくなります。ColPaliは、画像の上に高密度のベクトルのグリッドを作成するため、この作業に役立ちます。

上記の画像から、ColPaliが画像の特定の領域でAmazon BedrockとElasticがテキストとロゴの両方の形式で多く使用されていることを特定したことがわかります。Elasticのロゴと参照は実際にはかなり小さいですが、画像全体の高密度ベクトルとColPaliランクベクトルを使用した画像のリランキングの組み合わせにより、クエリでこのスライドを見つけて表示することができます。Elasticはこれらの機能を製品に組み込むことで、このプロセスを簡素化し、開発者が組織のニーズを解決するアプリケーションを簡単に作成できるようにしました。

公共部門におけるオープンで拡張性のあるベクトル検索とRAGのためのソリューション、Elastic

- 相互運用可能でLLMに依存しないサポート
- RAGアプローチでLLMトレーニングと微調整の必要性を排除
- アプリケーションスタックのツールとの完全なカスタマイズと統合を可能にする、十分に文書化されたAPI主導のアーキテクチャ
- 統合されたロールベースおよび属性ベースのセキュリティ制御（RBACおよびABAC）
- 世界で最もダウンロードされたベクトルデータベース
- あらゆる環境でAIサービスを登録するための推論API
- すぐに使用できる2つのベクトルモデル（E5は高密度エンコード用、ELSERは低密度エンコード用）と任意の他の埋め込みモデルを使用する機能
- ランキング品質、コンピューティングリソース、クエリレイテンシの削減に最適化されたベクトル検索機能
- より優れたバイナリ量子化（BBQ）でベクトルを圧縮し、検索と処理を高速化
- 高度な検索モデルとベクトル検索機能を活用した、エージェント型AIのための拡張性のある基盤

 © 2025 Elasticsearch B.V. All rights reserved.

elastic

次のステップに進む

ベクトル検索で組織のデータを活用する方法のデモをご覧になるには、公共部門の技術チームにお問い合わせください。elastic.co/jp/contact/public-sector

