

# AI でビジネスを拡大する

Red Hat\* OpenShift\* Data Science とインテル® アーキテクチャーを使用してオープンソースの AI テクノロジーを活用

Ted Jones Red Hat シニア・ソリューション・アーキテクト

Karl Eklund Red Hat 主席アーキテクト

Paulina Olszewska インテル コーポレーション クラウド・ソリューション・エンジニア

Piotr Grabuszynski インテル コーポレーション クラウド・ソリューション・エンジニア

AI は、未来のテクノロジーではなく、すでに現実のものとなっています。小売、銀行と金融、セキュリティ、ガバナンス、医療、産業は、AI が日常的に導入されている分野のほんの一部です。AI が生成したコンテンツと人間が生成したコンテンツの見分けがつかなくなる時代がそこまでやってきています。自然言語処理 (NLP) は、機械と人間の間の対話から 2 人の人間の間の対話に近いレベルまで進化しています。この進化により、政府、起業家、金融および医療分野の企業に新たな機会が訪れています。

さまざまなマシンラーニング (ML) タスクが容易になるように、インテルと Red Hat は新しいソリューションを開発しています。データマイニング、モデルのデプロイ、トレーニング、プロセスの自動化、データ転送はすべて、Red Hat\* OpenShift\* Data Science などの製品により強化されます。パフォーマンス向上のため、インテルは、インテル® アドバンスド・マトリクス・エクステンション (インテル® AMX) を含む新しいテクノロジーとアクセラレーターを提供しています。

## 問題ステートメント

AI は、さまざまな方法で活用できるビジネス上の利点を企業にもたらします。NLP は、大規模なテキストデータを効率良く解析することで、不正検出に役立ちます。感情分析を実行して、医療をサポートすることもできます。NLP により、より自然な形式で人間と機械の対話が可能になります。この記事では、企業に関する質問に答える AI モデルの簡単な使用法を示します。インテリジェント・アシスタントを使用すると、応答時間が短縮され、顧客対応に必要な人員の数が減り、コストが削減され、顧客の利便性が向上します。

## アプローチ

目標は、製品、価格、販売、企業に関する顧客の質問に答えることです。次の例は、オープンソースのコンポーネントを使用して、顧客と対話できるチャットボットをセットアップする簡単な方法を示しています。いくつかのプログラミングと開発を行うことで、ビジネスをサポートする、完全に機能する実用的なアプリケーションに拡張および変換できます。

この目標を達成するため、ここでは BERT (Bidirectional Encoder Representations from Transformers の略) とオープンソースの [BERT 質問応答 Python\\* デモ](#) (英語) を使用しています。次の例は、より迅速な応答を提供するため、Red Hat\* OpenShift\* Data Science プラットフォームと OpenVINO™ ツールキットの Operator を、インテル® AMX を搭載した第 4 世代インテル® Xeon® スケーラブル・プロセッサ上で実行しています。

## テクノロジー

### Red Hat\* OpenShift\* Data Science

[Red Hat\\* OpenShift\\* Data Science](#) (英語) は、インテリジェント・アプリケーションのデータ・サイエンティストやプログラマー向けのサービスであり、セルフマネージドまたはマネージド・クラウド・プラットフォームとして利用できます。ML モデルを実際の環境にデプロイする前に、迅速に開発、トレーニング、テストできる、完全にサポートされた環境を提供します。オンプレミス、パブリッククラウド、データセンター、エッジのいずれであっても、コンテナ上の本番環境で ML モデルをデプロイし、Red Hat\* OpenShift\* Data Science からほかのプラットフォームに簡単にエクスポートできます。

ML に Red Hat\* OpenShift\* Data Science を使用すると、多くの利点があります。このプラットフォームには、Jupyter Notebook\*、TensorFlow\*、PyTorch\*、OpenVINO™ など、データ・サイエンティストがワークフローで使用するさまざまな商用パートナーとオープンソースのツールおよびフレームワークが含まれています。Red Hat\* OpenShift\* Data Science は、安全でスケーラブルな環境を提供します。

## インテル® AMX

[インテル® AMX](#) は、CPU 上のディープラーニングのトレーニングと推論のパフォーマンスを向上させる新しい内蔵アクセラレーターです。[AI の処理が高速化され、パフォーマンスが最大 3 倍向上します](#)。自然言語処理、推奨システム、画像認識などのワークロードに最適です。第 4 世代インテル® Xeon® スケーラブル・プロセッサ上のパフォーマンスは、[インテル® oneAPI ベース・ツールキット](#)の一部であるインテル® oneAPI ディープ・ニューラル・ネットワーク・ライブラリー (インテル® oneDNN) を使用して微調整でき、TensorFlow\* と PyTorch\* AI フレームワーク、およびインテル® ディストリビューションの OpenVINO™ ツールキットに統合できます。これらのツールキットは Red Hat\* OpenShift\* Data Science で使用できます。

## 実装例

この例では、Super Shoes! という商店を説明する簡単なウェブサイトに基づいて使用します (図 1)。小売業者は、独自のウェブサイトを使用したり、会社に関する重要な情報を含む文書を作成することができます。次の例は、質問に回答できる簡単なチャットボットを開発し、顧客の観点から応答時間を短縮する方法を示しています。

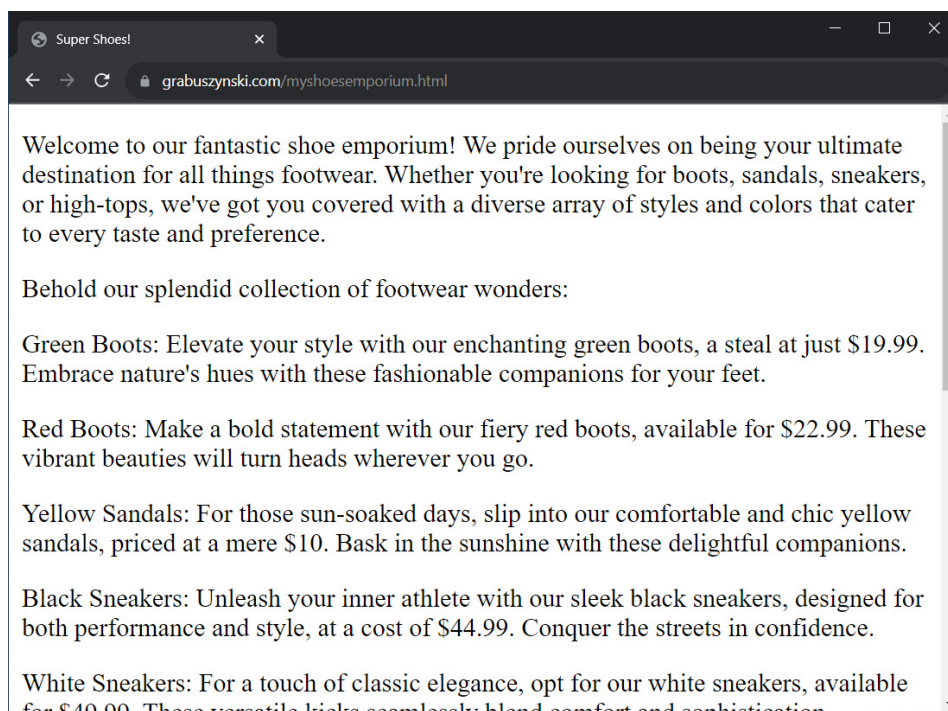


図 1. 例で使用したウェブサイトの一部

チャットボットの作成を開始するには、Red Hat\* OpenShift\* クラスターの[イメージ・レジストリーを準備します](#) (英語)。永続ボリューム (OpenShift\* Data Foundation など) を提供するストレージと、インテル® AMX をサポートする CPU を検出する Node Feature Discovery Operator も必要です。カーネルは実行時にインテル® AMX を検出するため、インテル® AMX を個別に有効にして構成する必要はありません。ノードのラベルをチェックして、インテル® AMX が利用可能であることを確認します。

```
feature.node.kubernetes.io/cpu-cpuid.AMXBF16=true
feature.node.kubernetes.io/cpu-cpuid.AMXINT8=true
feature.node.kubernetes.io/cpu-cpuid.AMX TILE=true
```

Red Hat\* OpenShift\* Data Science と OpenVINO™ ツールキットの Operator は、OpenShift\* Container Platform ウェブコンソールから利用できる OperatorHub からインストールする必要があります (図 2)。デフォルト設定で十分ですが、正しいインストールの順序 (最初に OpenShift\* Data Science、次に OpenVINO™ ツールキットの Operator) を守ってください。次に、redhat-ods-applications プロジェクトの [Installed Operators] タブに両方のオペレーターがリストされていること、ステータスが Succeeded になっていることを確認します。

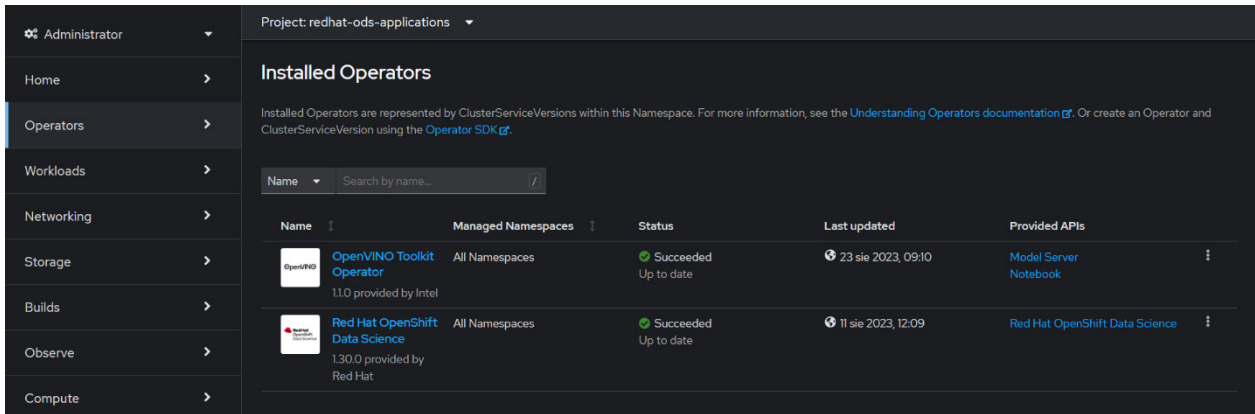


図 2. インストールされたオペレーターのステータスを示す OpenShift\* コンソールのスクリーンショット

OpenVINO™ ツールキットの Operator には追加の構成が必要です。redhat-ods-applications プロジェクトが選択されていることを確認します。OpenVINO™ ツールキットの Operator をクリックし、[Notebook] タブを選択します。デフォルト設定で新しいノートブックを作成します。これらの手順が完了したら、メインメニューから [Builds] > [Builds] に移動して、openvino-notebooks-v2022.3-1 ビルドが完了していることを確認します。

右上のメニューを使用して Red Hat\* OpenShift\* Data Science ダッシュボードに移動します (図 3)。

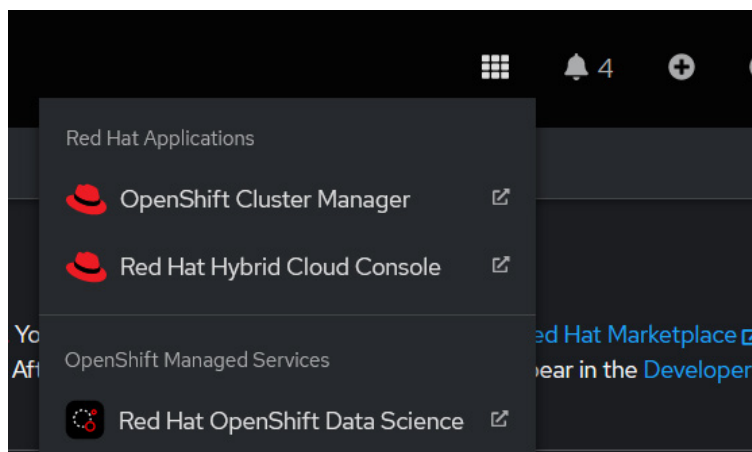


図 3. OpenShift\* Managed Services のメニュー

Red Hat\* OpenShift\* Data Science インターフェイスで、**[Applications]** > **[Enabled]** タブを選択します。OpenVINO™ ツールキット v2022.3 イメージを使用して Jupyter Notebook\* アプリケーションを起動します。コンテナのサイズを **[Large]** に変更します (図 4)。

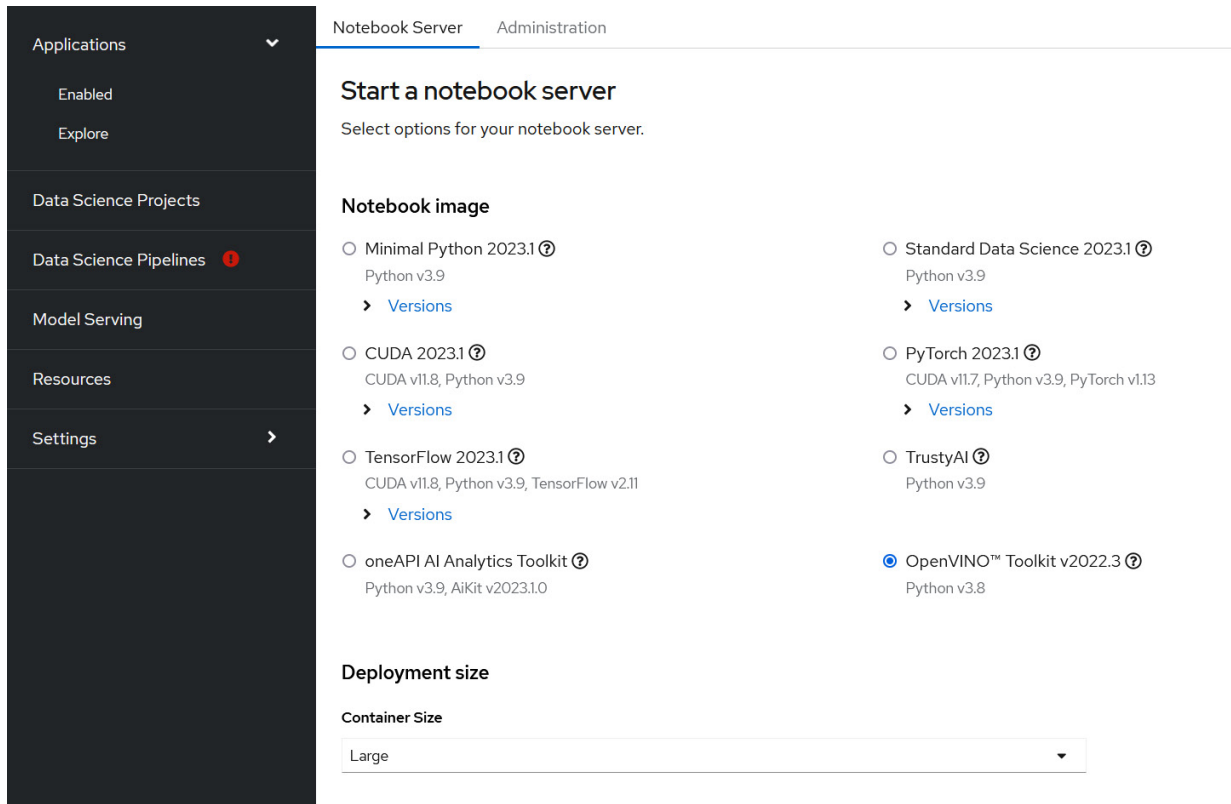


図 4. Red Hat\* OpenShift\* Data Science ダッシュボードで Jupyter Notebook\* を起動

ランチャーウィンドウで **[Terminal]** を選択します (図 5)。



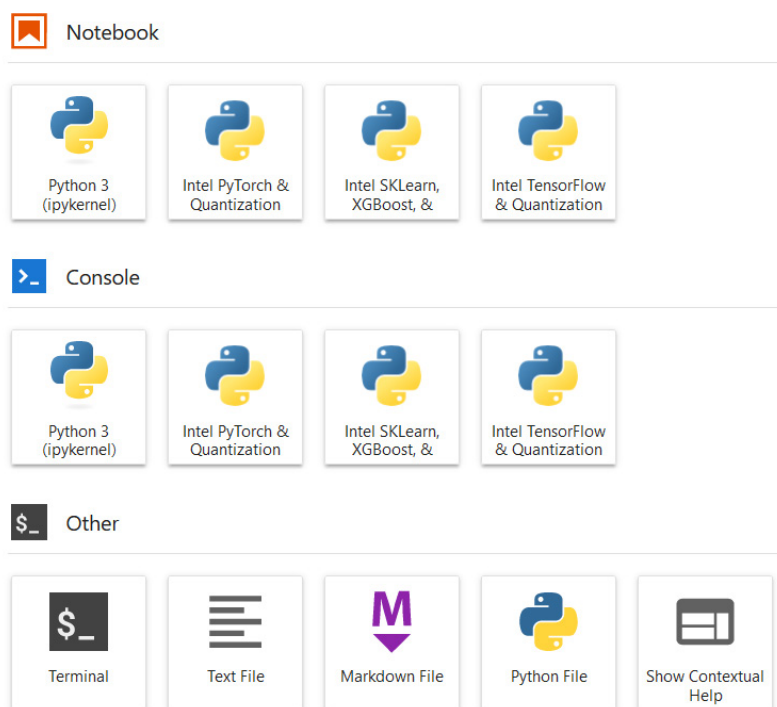


図 5. ランチャーウィンドウ

[demos](#) (英語) のソースコードを含む Open Model Zoo GitHub\* リポジトリをクローンします。

```
git clone --recurse-submodules
https://github.com/openvinotoolkit/open_model_zoo.git
```

コンパイルされたすべての命令を表示し、インテル® AMX の使用を制御するには、ONEDNN\_VERBOSE 環境変数を 1 に設定します。

```
export ONEDNN_VERBOSE=1
```

前述したように、この例は BERT 質問応答 Python\* デモをベースとしています。ただし、BERT-small の代わりに BERT-large を使用します。

open\_model\_zoo/demos/bert\_question\_answering\_demo/python ディレクトリーには、デモでサポートされているモデルのリストを含む models.lst ファイルがあります。モデルをダウンロードするには、次のコマンドでこのファイルを指定します。

```
omz_downloader --list models.lst
```

Red Hat\* OpenShift\* Data Science を使用する最も重要な利点の 1 つは、必要なツール（上記で使った [モデル・ダウンローダー](#) (英語) など) と Python\* パッケージがすべて組込まれていることです。OpenVINO™ ツールキットの Operator のおかげで、コンポーネントのインストールおよび開発環境の準備という、長く退屈でエラーが発生しやすいプロセスをスキップできます。

モデルがダウンロードできたら、最初のチャットボットを実行する準備は完了です。スクリプトを呼び出してみましょう。

```
python3 bert_question_answering_demo.py --vocab=/opt/app-root/src/open_model_zoo/
demos/bert_question_answering_demo/python/intel/bert-large-uncased-whole-word-
masking-squad-int8-0001/vocab.txt --model=/opt/app-root/src/open_model_zoo/
demos/bert_question_answering_demo/python/intel/bert-large-uncased-whole-word-
masking-squad-int8-0001/FP32-INT8/bert-large-uncased-whole-word-masking-squad-
int8-0001.xml --input_names="input_ids,attention_mask,token_type_ids" --output_
names="output_s,output_e" --input="https://grabuszynski.com/myshoesemporium.html" -c
```

サンプルページを入力として渡す代わりに、小売業者は自社のウェブサイトのサンプルを使用できます。現在、チャットボットは会社とその製品に関する質問に答えることができます (図 6)。

ONEDNN\_VERBOSE を 1 に設定した場合、インテル® AMX 命令が使用されていることを確認する avx\_512\_core\_amx がログに表示されます。

```

Type a question (empty string to exit): What is the price of green boots?
[ INFO ] Show top 3 answers
[ INFO ] Answer: $19.99
Score: 0.98
Context: Green Boots: Elevate your style with our enchanting green boots, a steal at just $19.99
[ INFO ] Answer: $49.99
Score: 0.03
Context: Hiking Boots Sale: Calling all adventure enthusiasts! Our robust hiking boots are now available at
a special price of $49.99

Type a question (empty string to exit): What are the opening hours?
[ INFO ] Show top 3 answers
[ INFO ] Answer: 8AM and remain wide open until 9PM
Score: 0.64
Context: Our doors open at 8AM and remain wide open until 9PM
[ INFO ] Answer: Welcome to our fantastic shoe emporium!
Score: 0.00
Context: Welcome to our fantastic shoe emporium! We pride ourselves on being your ultimate destination
for all things footwear

Type a question (empty string to exit): 
```

図 6. 質問の例

```
onednn_verbose,info,cpu,isa:Intel AVX-512 with float16, Intel DL Boost and bfloat16
support and Intel AMX with bfloat16 and 8-bit integer support
```

この例を拡張することもできます。モデルはモデルサーバーを使用して提供できます。使いやすくユーザーフレンドリーなインターフェイスを提供するアプリケーションは、このプロジェクトの特長となるでしょう。

## まとめ

Red Hat とインテルは、AI の新しいソリューション、テクノロジー、改善に常に取り組んでいます。この記事では、Red Hat\* OpenShift\* Data Science と OpenVINO™ ツールキットおよびインテル® AMX アクセラレーションを組み合わせることで、小売業界をどのように変えることができるかを説明しました。これらの最先端のテクノロジーをチャットボット・ソリューションで BERT モデルとともに使用することで、小売業者は次のことが可能になります。

- カスタマーサービスのコストを削減する
- 応答時間の短縮により顧客満足度を向上させる
- 潜在的に売上を伸ばす

Red Hat\* OpenShift\* Data Science とインテル® AMX アクセラレーションの強力な組み合わせにより、小売業者は大量のデータを迅速に処理して解析することができます。その結果として、より適切な意思決定と最適化された運用が可能になります。最終的に、この技術の融合により、小売業者は急速に変化する市場環境で競争力を維持するために必要なリソースを得ることができます。