







IBMのエンタープライズ向けAIとデータのプラットフォーム「watsonx.ai」を支える技術のご紹介

May 16, 2023





【米国ニューヨーク州アーモンク-2023年5月9日（現地時間）発】

著者： [デービッド・コックス \(David Cox\)](#) 、ITアーキテクト

私たちは、[AI革命の最前線](#)にいます。過去10年間で、ディープラーニングは、大量のデータと膨大な計算能力の融合により、多くの素晴らしいAI機能を可能にしました。しかし、私たちは、自動化には労力がかかるという逆説的な課題に直面しました。AIでビジネスの課題を解決しようとしたことがある人ならおわかりと思います。

日本IBM 広報代表
電話: 03-3808-5120
e-mail: PRESSREL@jp.ibm.com
[その他お問い合わせ窓口一覧](#)

Subscribe to email

従来のAIツールは強力ですが、高価で時間がかかり、使いこなすのが困難場合があります。AIモデルを学習させるためには、苦勞してデータを収集し、キュレーションし、タスクに特化した注釈を付けてラベル付けする必要があります。また、モデルを構築するためには、専門的なスキルが必要で、そのスキル習得が難しく、新しいタスクのたびに、そのプロセスを繰り返す必要があります。その結果、企業はデータが豊富でビジネス価値が高いタスクの自動化に主眼を置き、それ以外は棚上げしてきました。しかし、この状況は変わり始めています。

トランスフォーマーと自己教師あり学習法の登場により、膨大な量のラベルなしデータを利用できるようになり、"基盤モデル"と呼ばれる大規模な事前学習済みモデルへの道が開かれました。これらの大規模なモデルは、自動化に関わるコストと労力を低減させました。

基盤モデルは、様々なAIアプリケーションのための強力かつ汎用的な基盤を提供します。基盤モデルを使用すると、限られた注釈付きデータと最小限の労力でタスクを迅速に実行でき、場合によっては、ある課題を説明するだけで、モデルにその課題を解決させることができます。

しかし、これらの強力なテクノロジーは、企業にとって新たなリスクや課題ももたらします。今日のモデルの多くは、品質や出所が不明なデータセットで学習しているため、攻撃的、あるいはバイアスを含んだり、事実と異なったことを回答する場合があります。また、大規模なモデルは、学習と実行にかかるコストが高く、エネルギーを大量に消費し、導入も複雑です。

私たちIBMは、[エンタープライズ向け](#)の基盤モデルを利用するための本質的な課題を解決するアプローチを開発してきました。IBMは本日、市場にある最新のAIツールやテクノロジーにアクセスするための入り口としてIBMが提供する[watsonx.aiを発表](#)しました。この分野がいかに速く動いているかを物語るように、いくつかのツールはほんの数週間前に利用可能になったもので、この記事を書いている間にも新しいものが追加されています。

今回発表の[IBM watsonx](#)の製品群のひとつであるwatsonx.aiには様々な機能が含まれ、今後も進化し続けますが、私たちの最も重要なお約束は、企業向けの安全な自動化製品を提供するということです。

これは、AIにおける新しいパラダイムから価値を引き出すお客様の歩みを加速させるためのIBMの役割の一部です。ここでは、データおよびモデル・アーキテクチャーへのアプローチを含め、企業で活用できる、IBMで学習済みの一連の基盤モデルを構築してきた作業について説明します。また、私たちの新しいプラットフォームとツールについても紹介します。このプラットフォームでは、IBMが開発したモデルに加え、オープンソースの幅広いモデル・カタログを使用し、企業が基盤モデルを活用したソリューションを構築・展開することができます。

基盤モデルの土台となるデータ

データの品質は重要です。バイアスを含んだデータや有害なデータで学習されたAIモデルは、当然、バイアスを含んだ回答や有害なアウトプットを生成する傾向があります。この問題は、基盤モデルの時代になるとさらに悪化します。基盤モデルでは、モデルの学習に使用するデータは通常、多くのソースから取得され、人間がすべてを十分精査することができないほど大量になるからです。

データは基盤モデルを動かす燃料であるため、IBMは、モデルに入るすべてのものを綿密にキュレーションすることに力を注いできました。私たちは、ヘイトスピーチや下品な言葉、ライセンスによる制限、偏見（バイアス）がないか、データを徹底的にフィルタリングするAIツールを開発しました。好ましくないデータが特定された場合は、それを削除し、モデルを再トレーニングすることを繰り返します。

データのキュレーションは、終わることのない作業です。IBMは、データの品質と管理を向上させ、改正される法律や規制の要件を満たすために、新しい手法を開発し、改良続けています。私たちは、クリーニングする前の元データ、使用した手法、そして各データ・ポイントを利用したモデルのそれぞれを追跡するエンドツーエンドのフレームワークを構築しています。

金融、法律、サイバーセキュリティ、サステナビリティなど、さまざまな領域で最も差し迫ったビジネス課題に取り組むために、私たちは高品質のデータを収集し続けています。現在、基盤モデルの学習用に1テラバイト以上のキュレーションされたテキストの収集を目標としていますが、さらにキュレーションされたソフトウェアコード、衛星データ、ITネットワークのイベント・データおよびログも追加しています。

IBM Research は、基盤モデルのライフサイクルを通じて信頼を担保し、バイアスを緩和してモデルの安全性を向上させる技術も開発しています。この分野での私たちの研究には、モデルの調整に使用するデータからバイアスを含んだデータ・ポイントを特定し、それを編集できるようにする[FairIJ](#)があります。また、[フェアネス・リプログラミング](#)のように、学習後のモデルでもバイアスを軽減することができる手法もあります。

金融、法律、サイバーセキュリティ、サステナビリティなど、さまざまな領域で最も差し迫ったビジネス課題に取り組むために、私たちは高品質のデータを収集し続けています。現在、基盤モデルの学習用に1テラバイト以上のキュレーションされたテキストの収集を目標としていますが、さらにキュレーションされたソフトウェアコード、衛星データ、ITネットワークのイベント・データおよびログも追加しています。

IBM Research は、基盤モデルのライフサイクルを通じて信頼を担保し、バイアスを緩和してモデルの安全性を向上させる技術も開発しています。この分野での私たちの研究には、モデルの調整に使用するデータからバイアスを含んだデータ・ポイントを特定し、それを編集できるようにする[FairIJ](#)があります。また、[フェアネス・リプログラミング](#)のように、学習後のモデルでもバイアスを軽減することができる手法もあります。

企業価値重視の効率的な基盤モデル

IBMの新しいwatsonx.aiのスタジオ（ツール・機能群）は、企業価値の提供を目的とした[一連の基盤モデル](#)を提供します。これらは様々なIBM製品に組み込まれ、今後数ヶ月でIBMのお客様に提供していく予定です。

1つの基盤モデルがすべてのニーズに対応できるわけではないという認識に基づいて、私たちはさまざまな規模とアーキテクチャーで、自然言語とプログラム言語の基盤モデル・ファミリーを構築しています。各モデル・ファミリーには、Granite（花崗岩）、Sandstone（砂岩）、Obsidian（黒曜石）、Slate（粘板岩）といった地質学をテーマにしたコードネームが付いており、IBM Research とオープン・リサーチ・コミュニティからの最先端のイノベーションを結集しています。各モデルは、企業のさまざまなタスクに合わせてカスタマイズすることができます。

Graniteモデルは、デコーダーのみのGPTに似たアーキテクチャーを採用し、生成タスクに適しています。**Sandstone**モデルは、エンコーダー・デコーダー・アーキテクチャーを採用し、特定のタスクでの微調整に適し、よく使われているGoogleのモデルT5と互換性があります。**Obsidian**は、IBM Researchが開発した新しいモジュール型アーキテクチャーを採用したモデルで、さまざまなタスクにおいて高い推論効率と性能レベルを提供します。**Slate**は、エンコーダーのみのRoBERTa ベースのモデル群を指し、生成型ではありませんが、多くの企業向けNLPタスクに高速かつ効果的に対応します。watsonx.aiのモデルはすべて、当社のカスタム設計のクラウド・ネイティブAIスーパーコンピュータ「[Vela](#)」上で、IBMが管理する企業向けのデータレイクを用いて学習されています。

効率性とサステナビリティは、watsonx.aiの中核となる設計の原則です。IBM Researchでは、小さなモデルを再利用して大きなモデルに「成長」させる「[LIGO](#)」アルゴリズムなど、効率的なモデル学習のための新しい技術を開発してきました。この方法によって、モデルの学習に必要な時間、コスト、二酸化炭素排出量を40%~70%削減することができます。推論速度を向上させるために、私たちは[量子化](#)に関する専門知識を活用し、モデルを 32 ビット浮動小数点フォーマットよりはるかに小さい整数ビットフォーマットに縮小しています。AIモデルの精度を下げることで、精度を犠牲にすることなく効率性が大幅に向上します。この圧縮したモデルを、AIに最適化されたチップである「[IBM AIU](#)」でもまもなく実行できるようにしたいと考えています。

Graniteモデルは、デコーダーのみのGPTに似たアーキテクチャーを採用し、生成タスクに適しています。**Sandstone**モデルは、エンコーダー・デコーダー・アーキテクチャーを採用し、特定のタスクでの微調整に適し、よく使われているGoogleのモデルT5と互換性があります。**Obsidian**は、IBM Researchが開発した新しいモジュール型アーキテクチャーを採用したモデルで、さまざまなタスクにおいて高い推論効率と性能レベルを提供します。**Slate**は、エンコーダーのみのRoBERTa ベースのモデル群を指し、生成型ではありませんが、多くの企業向けNLPタスクに高速かつ効果的に対応します。watsonx.aiのモデルはすべて、当社のカスタム設計のクラウド・ネイティブAIスーパーコンピュータ「[Vela](#)」上で、IBMが管理する企業向けのデータレイクを用いて学習されています。

効率性とサステナビリティは、watsonx.aiの中核となる設計の原則です。IBM Researchでは、小さなモデルを再利用して大きなモデルに「成長」させる「[LIGO](#)」アルゴリズムなど、効率的なモデル学習のための新しい技術を開発してきました。この方法によって、モデルの学習に必要な時間、コスト、二酸化炭素排出量を40%~70%削減することができます。推論速度を向上させるために、私たちは[量子化](#)に関する専門知識を活用し、モデルを 32 ビット浮動小数点フォーマットよりはるかに小さい整数ビットフォーマットに縮小しています。AIモデルの精度を下げることで、精度を犠牲にすることなく効率性が大幅に向上します。この圧縮したモデルを、AIに最適化されたチップである「[IBM AIU](#)」でもまもなく実行できるようにしたいと考えています。

Graniteモデルは、デコーダーのみのGPTに似たアーキテクチャーを採用し、生成タスクに適しています。**Sandstone**モデルは、エンコーダー・デコーダー・アーキテクチャーを採用し、特定のタスクでの微調整に適し、よく使われているGoogleのモデルT5と互換性があります。**Obsidian**は、IBM Researchが開発した新しいモジュール型アーキテクチャーを採用したモデルで、さまざまなタスクにおいて高い推論効率と性能レベルを提供します。**Slate**は、エンコーダーのみのRoBERTa ベースのモデル群を指し、生成型ではありませんが、多くの企業向けNLPタスクに高速かつ効果的に対応します。watsonx.aiのモデルはすべて、当社のカスタム設計のクラウド・ネイティブAIスーパーコンピュータ「[Vela](#)」上で、IBMが管理する企業向けのデータレイクを用いて学習されています。

効率性とサステナビリティは、watsonx.aiの中核となる設計の原則です。IBM Researchでは、小さなモデルを再利用して大きなモデルに「成長」させる「[LIGO](#)」アルゴリズムなど、効率的なモデル学習のための新しい技術を開発してきました。この方法によって、モデルの学習に必要な時間、コスト、二酸化炭素排出量を40%~70%削減することができます。推論速度を向上させるために、私たちは[量子化](#)に関する専門知識を活用し、モデルを 32 ビット浮動小数点フォーマットよりはるかに小さい整数ビットフォーマットに縮小しています。AIモデルの精度を下げることで、精度を犠牲にすることなく効率性が大幅に向上します。この圧縮したモデルを、AIに最適化されたチップである「[IBM AIU](#)」でもまもなく実行できるようにしたいと考えています。

基盤モデル向けハイブリッドクラウド・ツール

基盤モデルには、モデルのチューニングとデプロイのための使いやすいソフトウェア・プラットフォームを作ることも必要です。RedHat OpenShift上に構築したIBMのハイブリッドでクラウド・ネイティブな[AIスタック](#)は、基盤モデルの学習および提供に最適化されています。企業は、OpenShiftの柔軟性を活用して、オンプレミスを含むあらゆる場所からモデルを実行することができます。

IBMは、基盤モデルに基づいたソリューションを構築するための、使いやすいユーザー・インターフェースと開発者向けライブラリーをwatsonx.aiのツール群として開発しました。Prompt Labでは、ラベルをつけた数個の例を与えるだけでAIタスクを迅速に実行することができます。Tuning Studioは、[IBM Researchが開発](#)した最先端の効率的なファイン・チューニング技術に基づき、お客様自身のデータを使ってモデルの迅速かつ堅牢なカスタマイズを可能にします。

IBMは、基盤モデルに基づいたソリューションを構築するための、使いやすいユーザー・インターフェースと開発者向けライブラリーをwatsonx.aiのツール群として開発しました。Prompt Labでは、ラベルをつけた数個の例を与えるだけでAIタスクを迅速に実行することができます。Tuning Studioは、[IBM Researchが開発](#)した最先端の効率的なファイン・チューニング技術に基づき、お客様自身のデータを使ってモデルの迅速かつ堅牢なカスタマイズを可能にします。

IBMは、基盤モデルに基づいたソリューションを構築するための、使いやすいユーザー・インターフェースと開発者向けライブラリーをwatsonx.aiのツール群として開発しました。Prompt Labでは、ラベルをつけた数個の例を与えるだけでAIタスクを迅速に実行することができます。Tuning Studioは、[IBM Researchが開発](#)した最先端の効率的なファイン・チューニング技術に基づき、お客様自身のデータを使ってモデルの迅速かつ堅牢なカスタマイズを可能にします。

[watsonx.ai](#)は、IBM独自のモデルを提供するだけでなく、オープンソース・モデルを企業が実験し迅速に試行錯誤できるように幅広いモデルのカタログへのシームレスなアクセスを提供します。IBMは、Hugging Faceとの新たなパートナーシップにより、何千ものオープンソースHugging Face基盤モデル、データセット、ライブラリーを[watsonx.ai](#)内で提供します。一方、Hugging Faceは、[watsonx.ai](#)上のIBM独自およびオープン・アクセスのモデルやツールを提供します。

ドロップダウンメニューから選択するだけで新しいモデルをお試しいただけます。スタジオの詳細については、[こちら](#)をご覧ください。

未来への展望

基盤モデルは、AIの展望を変えつつあり、進歩はさらに加速しています。私たちIBMは、この急速に進化する分野の最前線でイノベーションを真の企業価値に変えることをご支援してまいります。

以上

当報道資料は、2023年5月9日（現地時間）にIBM Corporationが発表したブログの抄訳です。原文は[こちら](#)をご参照ください。

Release Categories

AI



United States — English



Products & Solutions

Top products & platforms
Industries
Artificial intelligence
Blockchain
Business operations
Cloud computing
Data & Analytics
Hybrid cloud
IT infrastructure
Security
Supply chain
Financing

Learn about

What is Hybrid Cloud?
What is Artificial Intelligence?
What is Cloud Computing?
What is Kubernetes?
What are Containers?
What is DevOps?
What is Machine Learning?

Popular links

IBM Consulting
Communities
Developer education
Support - Download fixes, updates & drivers
IBM Research
Partner with us - Partner Plus
Training - Courses
Upcoming events & webinars

About IBM

Annual report
Career opportunities
Corporate social responsibility
Diversity & inclusion
Industry analyst reports
Investor relations
News & announcements
Thought leadership
Security, privacy & trust
About IBM

Follow IBM

LinkedIn
Twitter
Instagram
Subscription Center

Contact IBM

Privacy

Terms of use

Accessibility

Cookieの設定