



機械学習モデル開発プロジェクトの進め方

ML Enablement Series: Introduction

AWS Black Belt Online Seminar

久保 隆宏

Developer Relation
Machine Learning

© 2022, Amazon Web Services, Inc. or its Affiliates.

AWS 公式 Webinar

<https://amzn.to/JPWebinar>



過去資料

<https://amzn.to/JPArchive>



本動画で解説すること3つ

1. 機械学習モデルの開発プロセスと組織
2. 機械学習を使うメリット・デメリット
3. 開発プロセスを解説する流れ

本動画はシリーズ1本目です

想定する視聴者

1. **プログラミング経験があり機械学習に興味があるエンジニアの方**
2. 機械学習を活用するプロジェクトがはじまってしまったプロジェクトマネージャー、プロダクトマネージャーの方
3. 機械学習の活用プロジェクトに参画するデータサイエンティストの方

一般的な機械学習解説教材のイメージ



機械学習モデル

本動画シリーズのイメージ

プロダクト

機械学習モデル単体の開発でなく、
プロダクトでの活用方法を知りたい方にお勧め

本動画シリーズ全体で得られること

- Pythonで機械学習の基本的な実装を行えるようになる。
- 開発に関わる他のステークホルダーに、Jupyter Notebookを提示しながら必要な確認をとれるようになる。
- データサイエンティストに対し適切な声掛けができるようになる。

データサイエンティストは今まで開発チームにあまりいなかったロール。適切な接し方をしないとモチベーションを維持できない。

参考: [データサイエンティストの育て方](#)

参考: [StackOverflow 2020年の調査](#): 20%のデータサイエンティストは転職先を探している。

本動画シリーズのゴール

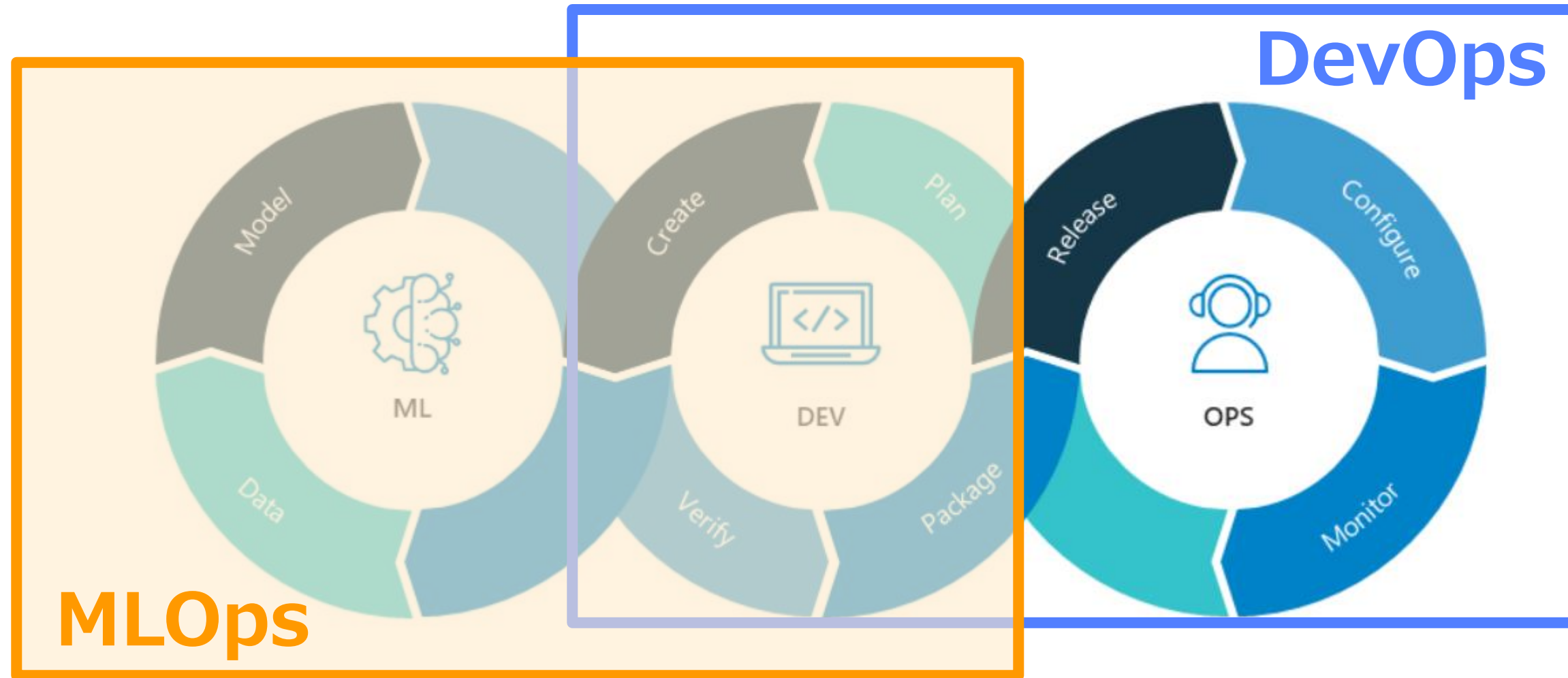
データサイエンティストを含むプロダクト開発チームが、機械学習モデルの開発において足りないスキルを補い合い、顧客に価値を届けられるようになること。



本日解説すること

- 1.機械学習モデルの開発プロセスと組織
- 2.機械学習を使うメリット・デメリット
- 3.開発プロセスを解説する流れ

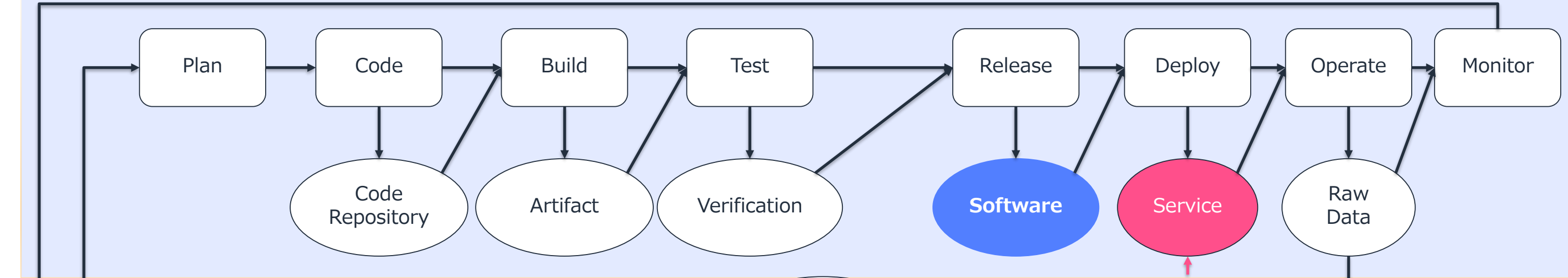
機械学習を利用するプロダクトでは、ソフトウェア開発のDevOpsに機械学習モデル開発のMLOpsが加わる。



出典: [Neal Analytics](#)

機械学習モデルの開発が加わったイメージ。

DevOps

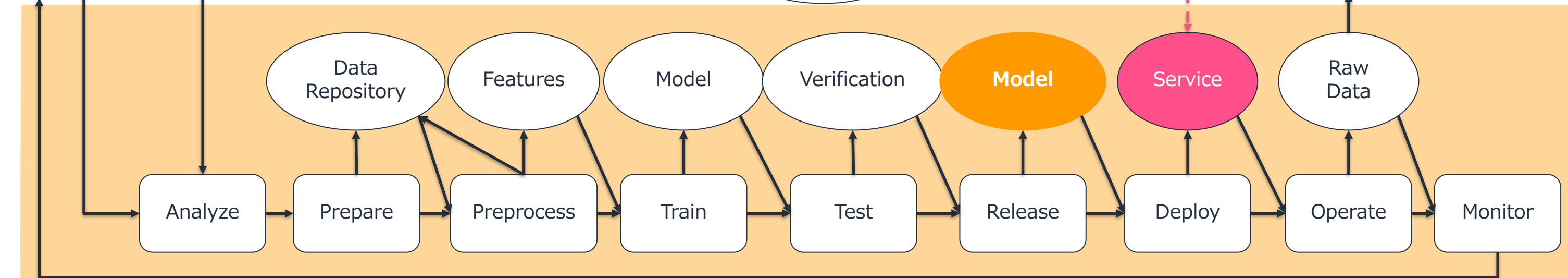


Business Understanding

Data Warehouse

Collect

MLOps



Analyze

Prepare

Preprocess

Train

Test

Release

Deploy

Operate

Monitor

Data Repository

Features

Model

Verification

Model

Service

Raw Data

DevOpsのおさらい



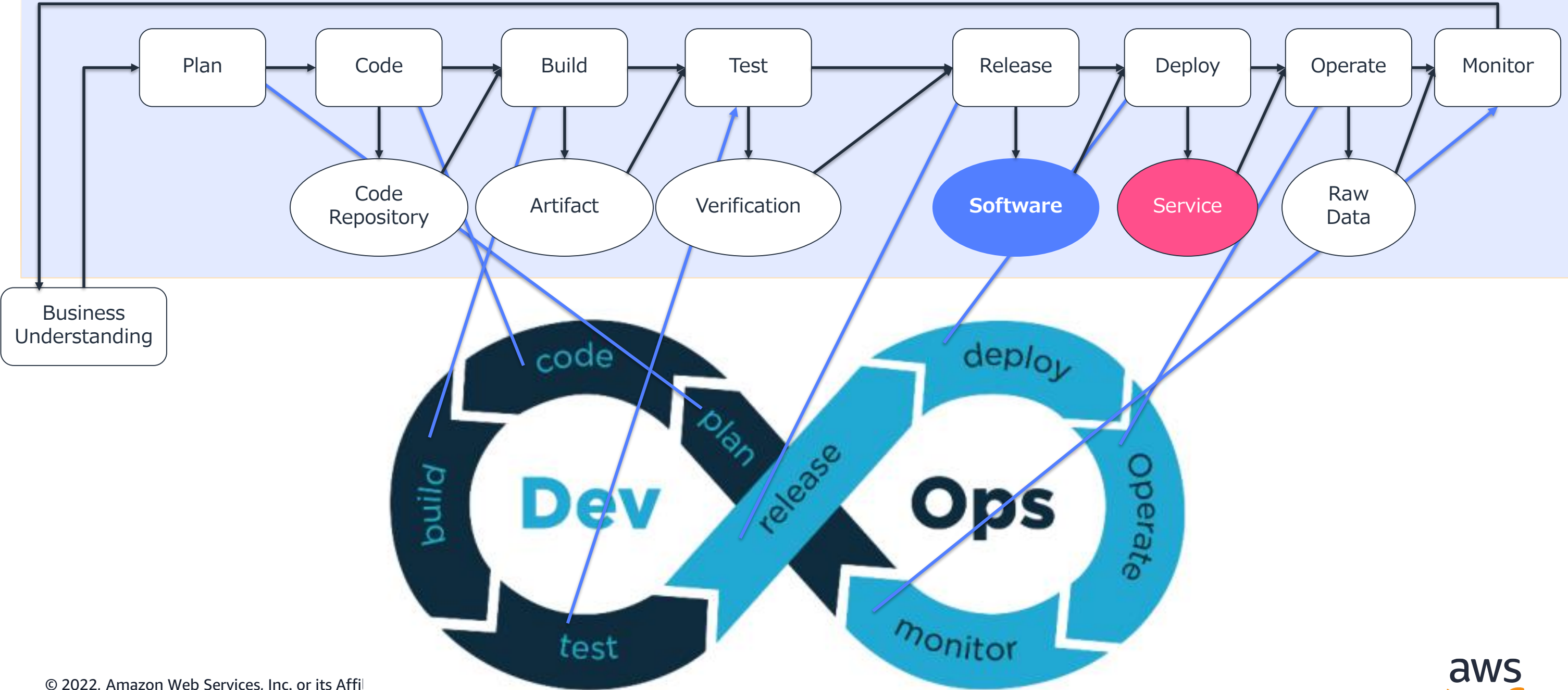
出典: [Neal Analytics](#)

DevOpsは開発（Development）と運用（Operations）を組み合わせた造語です。分断されていた開発チームと運用チームが協調させる。

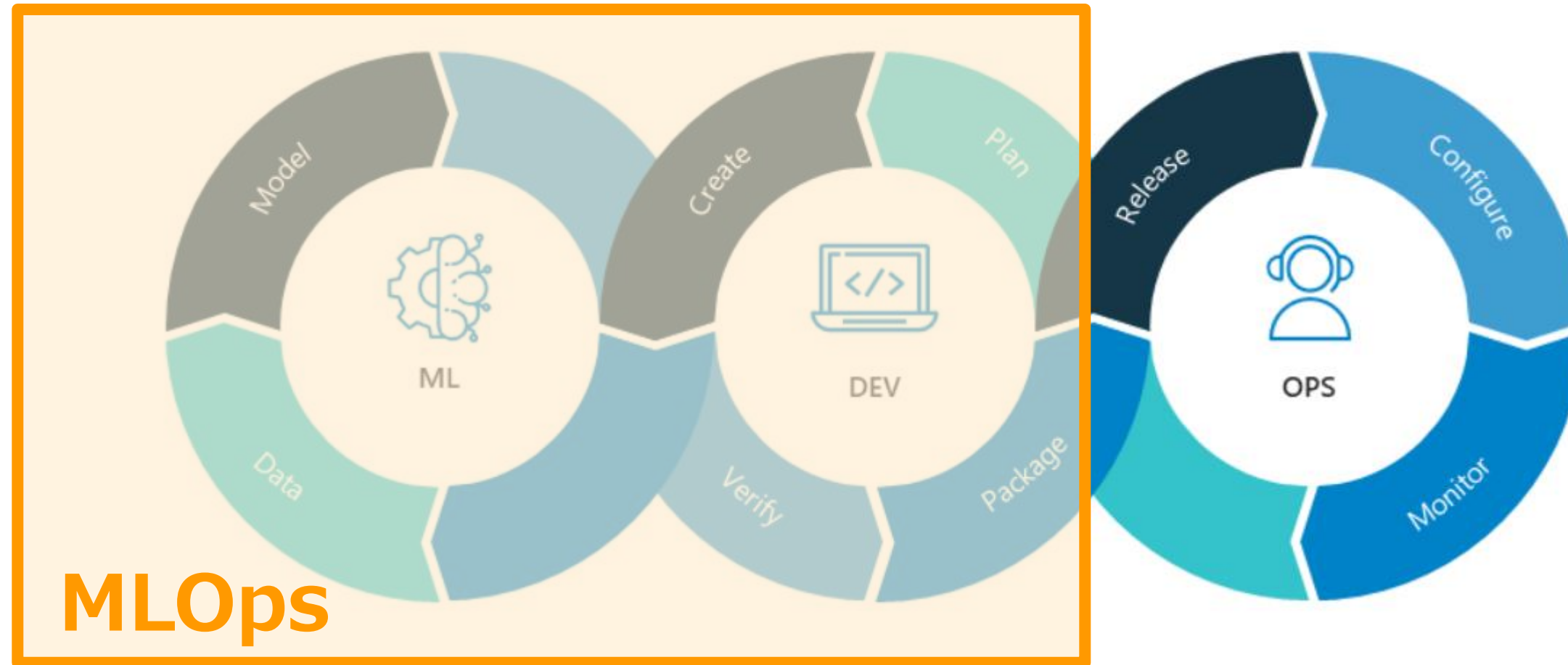
1. ソフトウェア／システムのビジネス価値をより高める
2. ビジネスの価値をより確実かつ迅速にエンドユーザーに届け続ける

上のパートはDevOpsのプロセス

DevOps

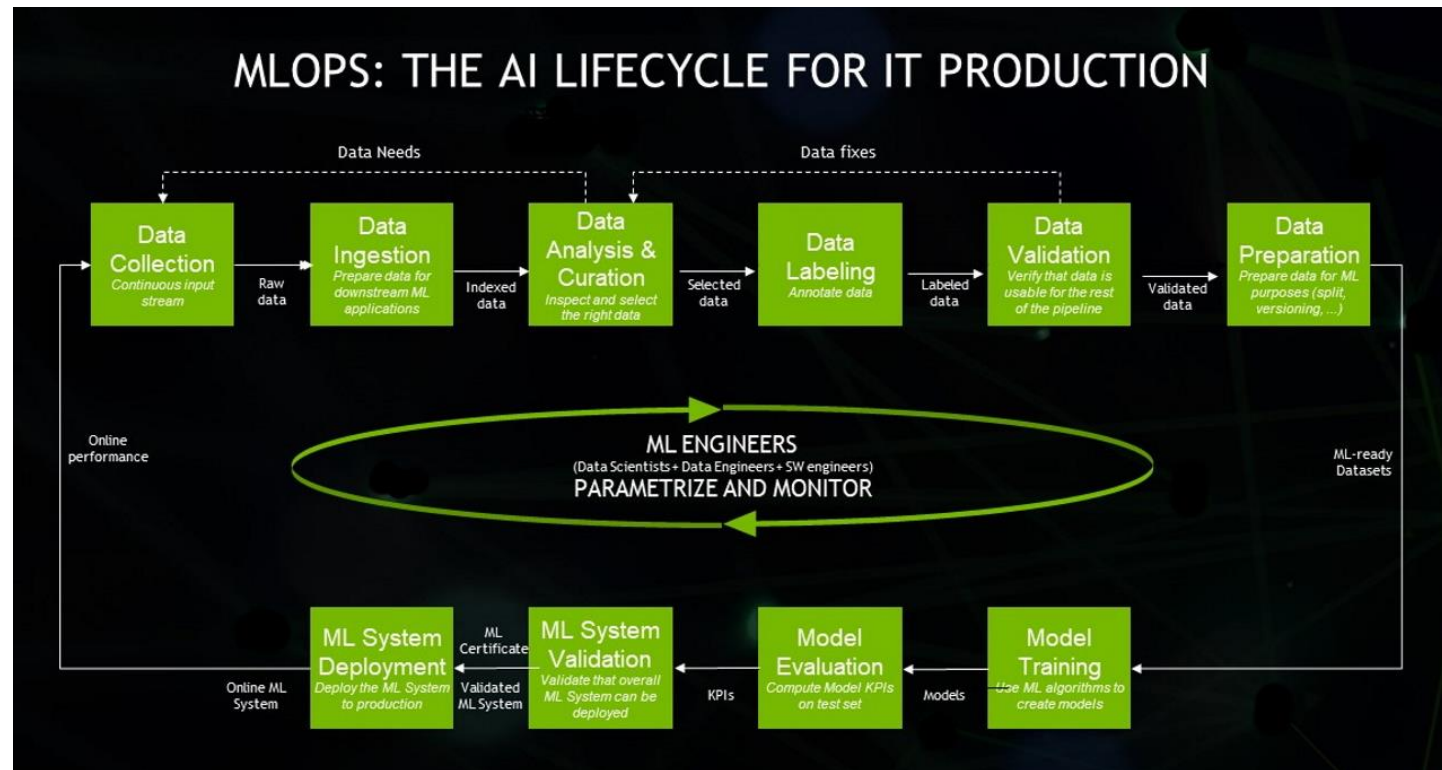


MLOps は、DevOps の開発手法や組織文化を、 機械学習の開発に取り込んだ開発手法

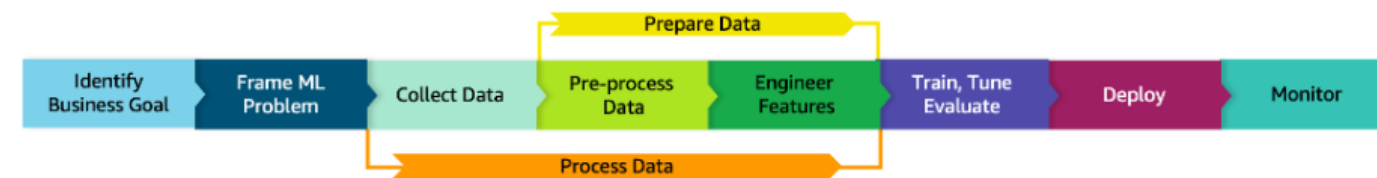


出典: [Neal Analytics](#)

MLOpsのプロセス定義はまだ固まったものはないが . . .

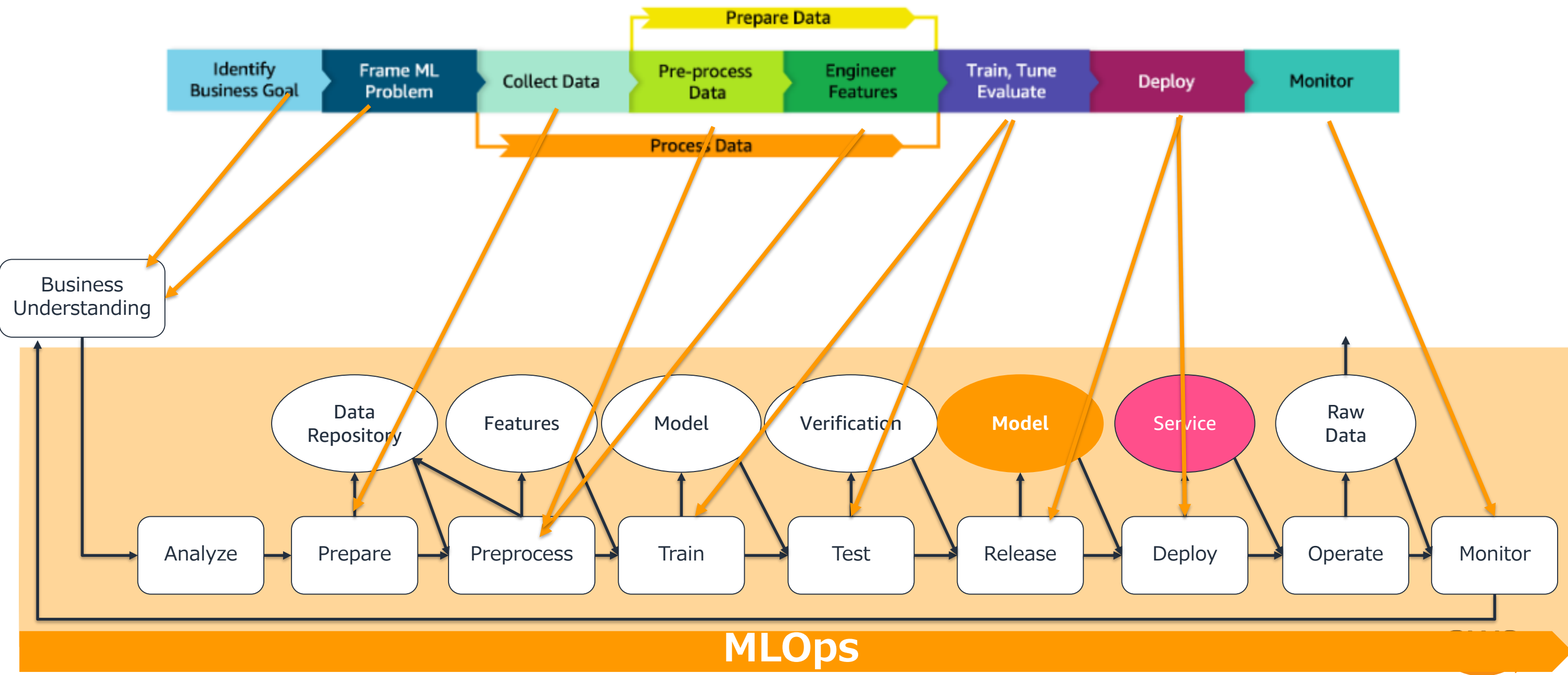


出典: NVIDIA



出典: AWS(ML Lens)

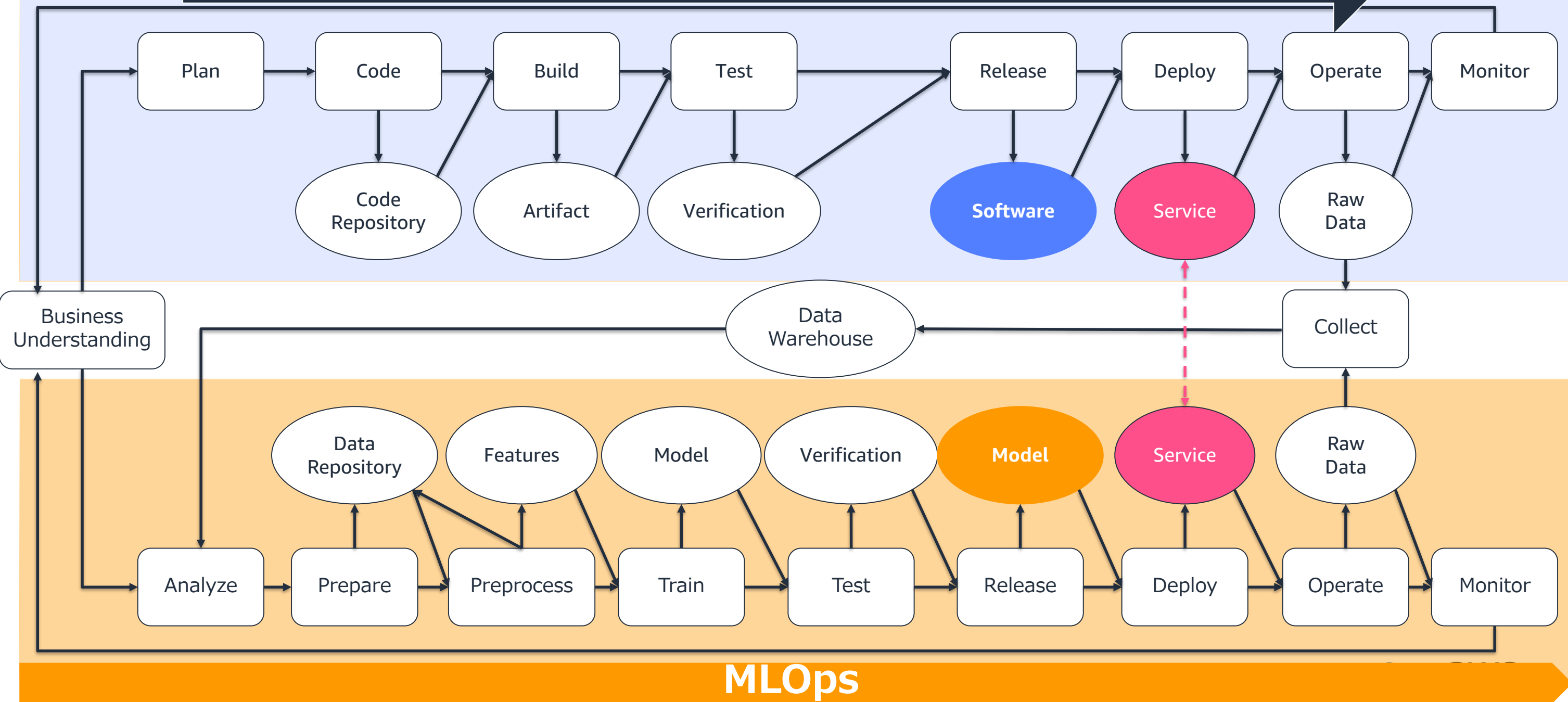
DevOpsのプロセス名と対応するよう、主にAWSの定義 (MLLens)を参照し定義しています。



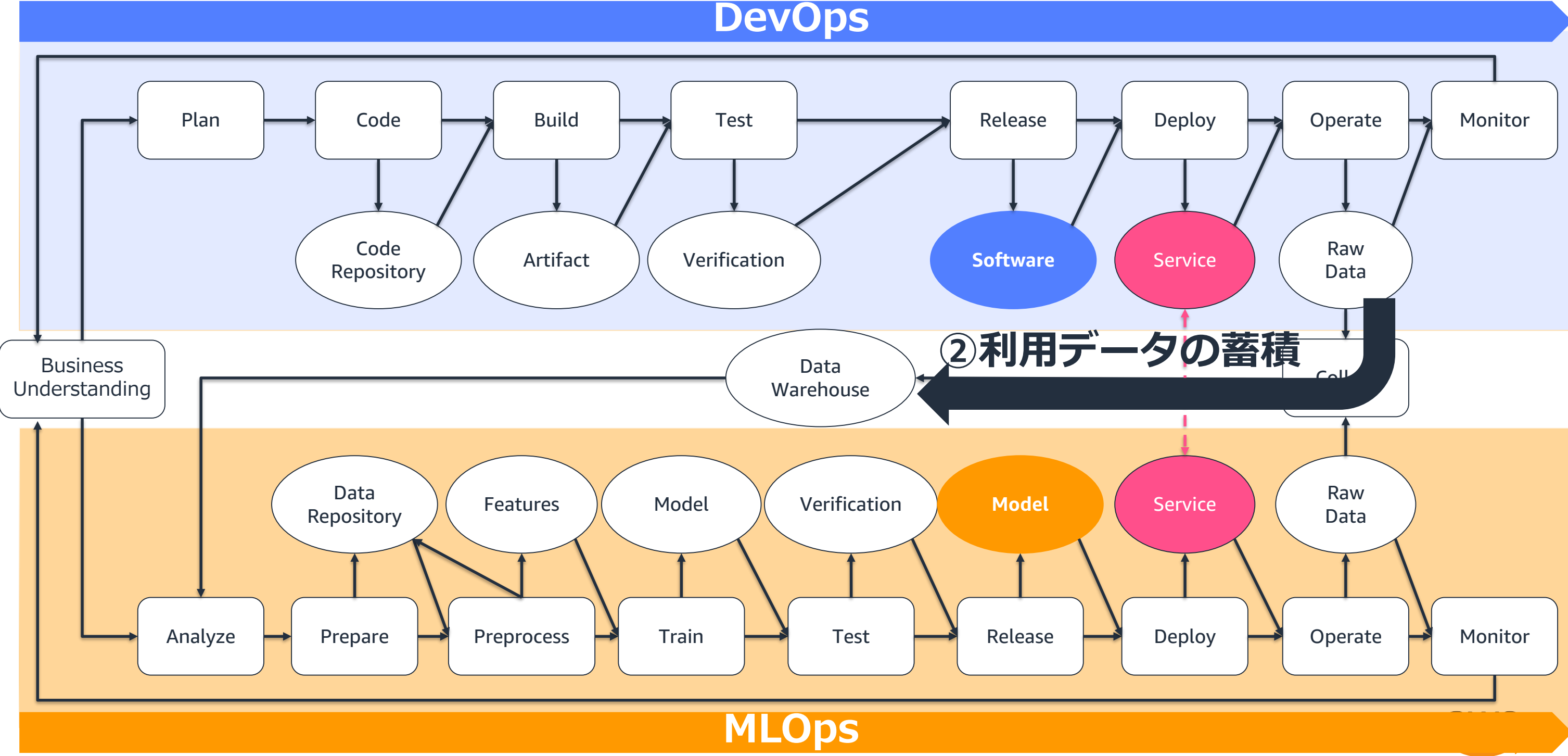
開発の流れのイメージをつかもう！

DevOpsとMLOpsの並走する開発のイメージ

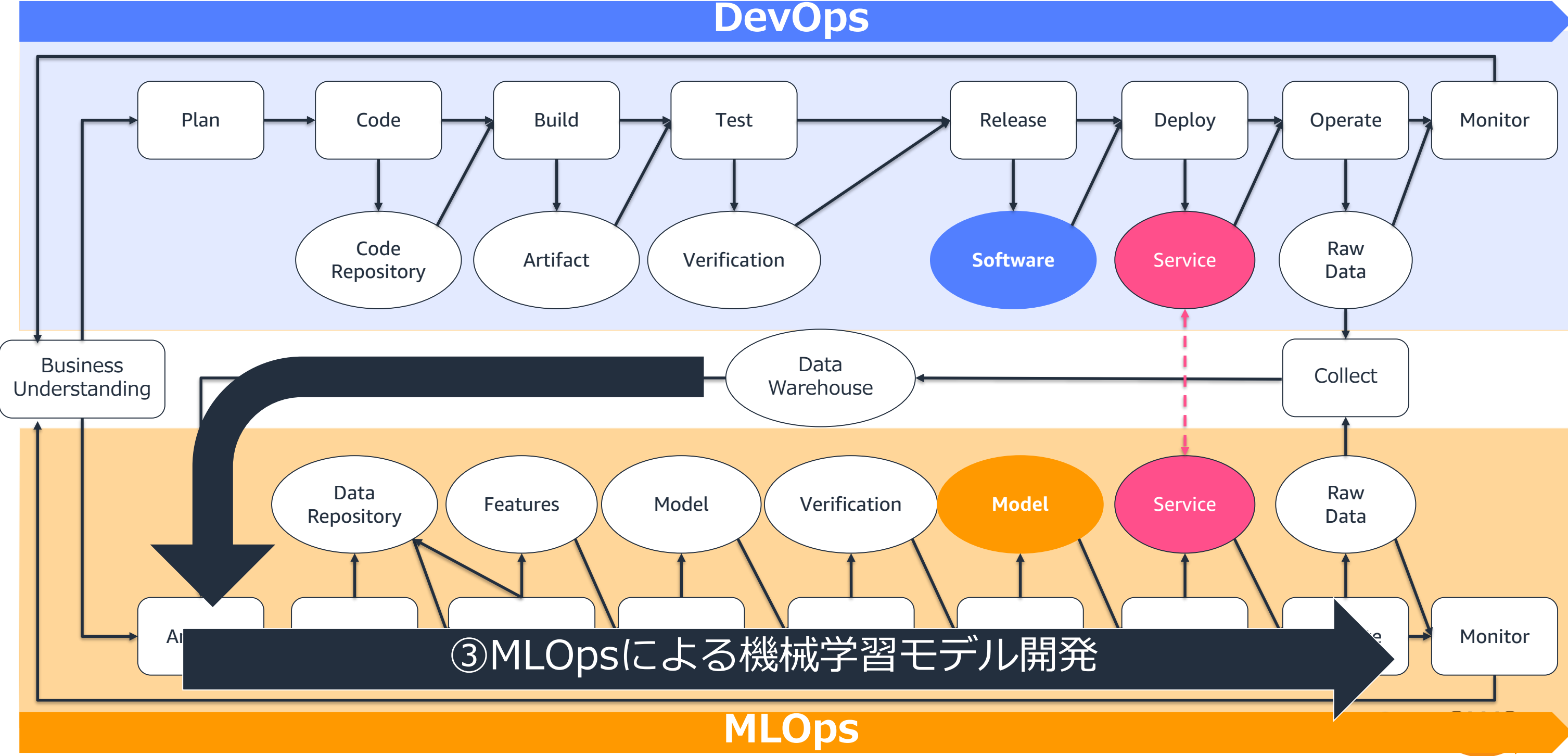
① DevOpsによるソフトウェア開発



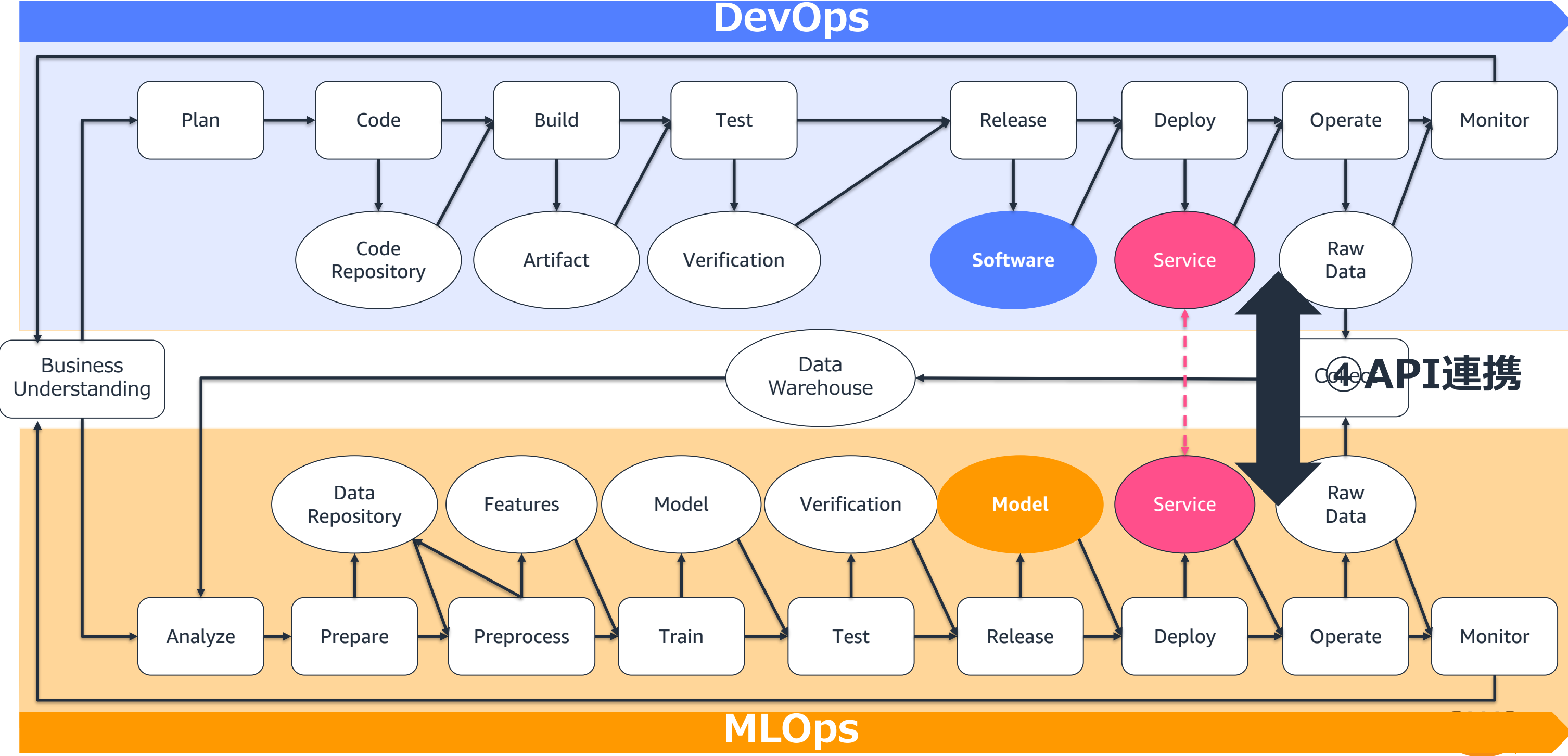
DevOpsとMLOpsの並走する開発のイメージ



DevOpsとMLOpsの並走する開発のイメージ

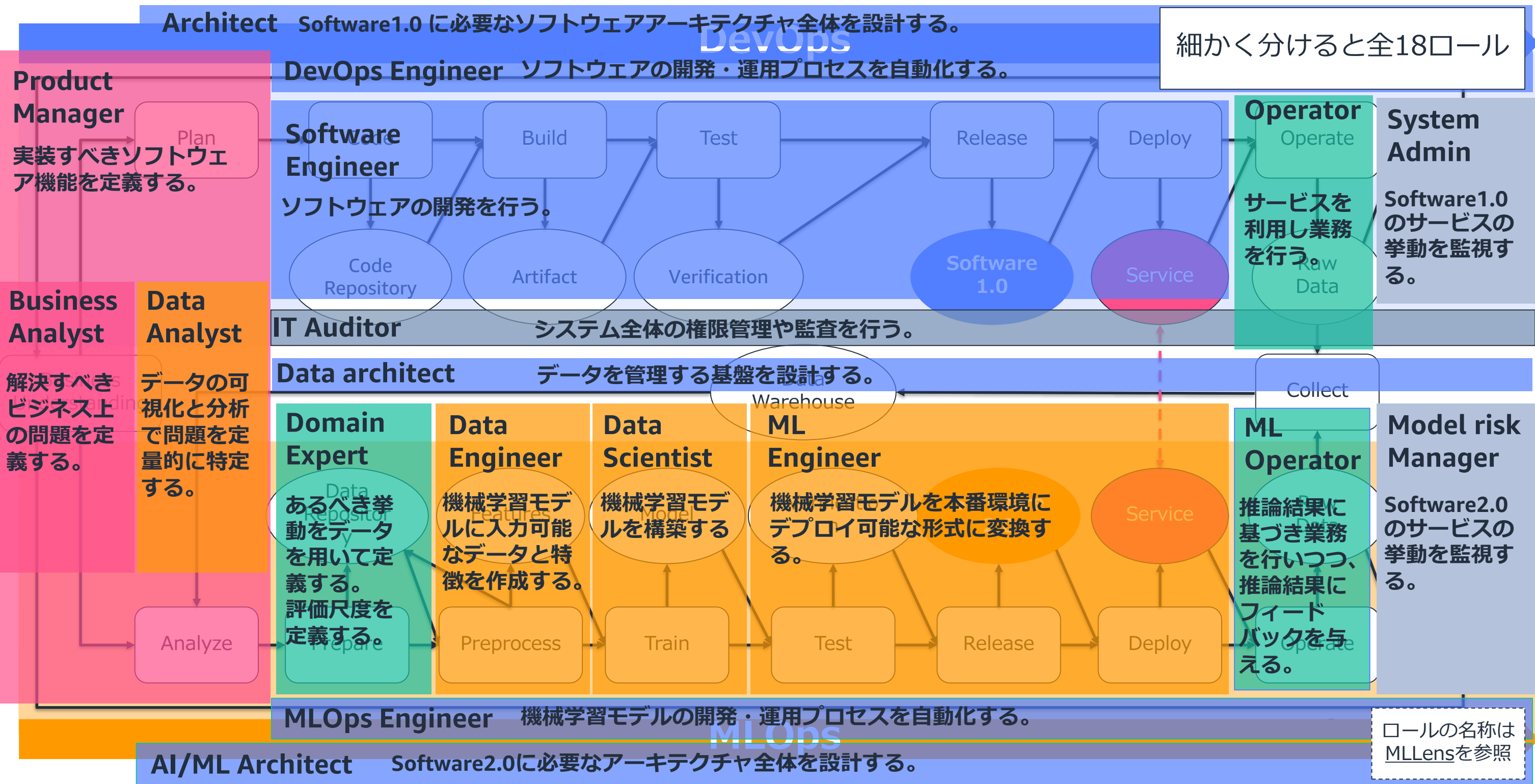


DevOpsとMLOpsの並走する開発のイメージ



**開発プロセスはわかった。
では組織は？**

様々なロールの協力で機械学習モデルはプロダクトに配置される



データサイエンティストだけでは機械学習をリリースできない

Architect Software1.0に必要なソフトウェアアーキテクチャ全体を設計する。

Product Manager
実装すべきソフトウェア機能を定義する。

DevOps Engineer ソフトウェアの

Software Engineer

ソフトウェアの開発を行う。

Code Repository

Artifact

なんとなくデータサイエンティストに期待されている範囲は広すぎる。
(データサイエンティストとしては“Data Scientist”のパートだけやりたい)

Business Analyst
解決すべき

Data Analyst
データの可

ビジネス上の問題を定義する。

視化と分析で問題を定量的に特定する。

Analyze

IT Auditor

システム全体の権限管理や監査を行う。

Data architect

データを管理する基盤を設計する。

Domain Expert

あるべき挙動をデータを用いて定義する。評価尺度を定義する。

Data Engineer

機械学習モデルに入力可能なデータと特徴を作成する。

Data Scientist

機械学習モデルを構築する

Warehouse ML Engineer

機械学習モデルを本番環境にデプロイ可能な形式に変換する。

ML Operator

推論結果に基づき業務を行いつつ、推論結果にフィードバックを与える。

Model risk Manager

Software2.0のサービスの挙動を監視する。

MLOps Engineer

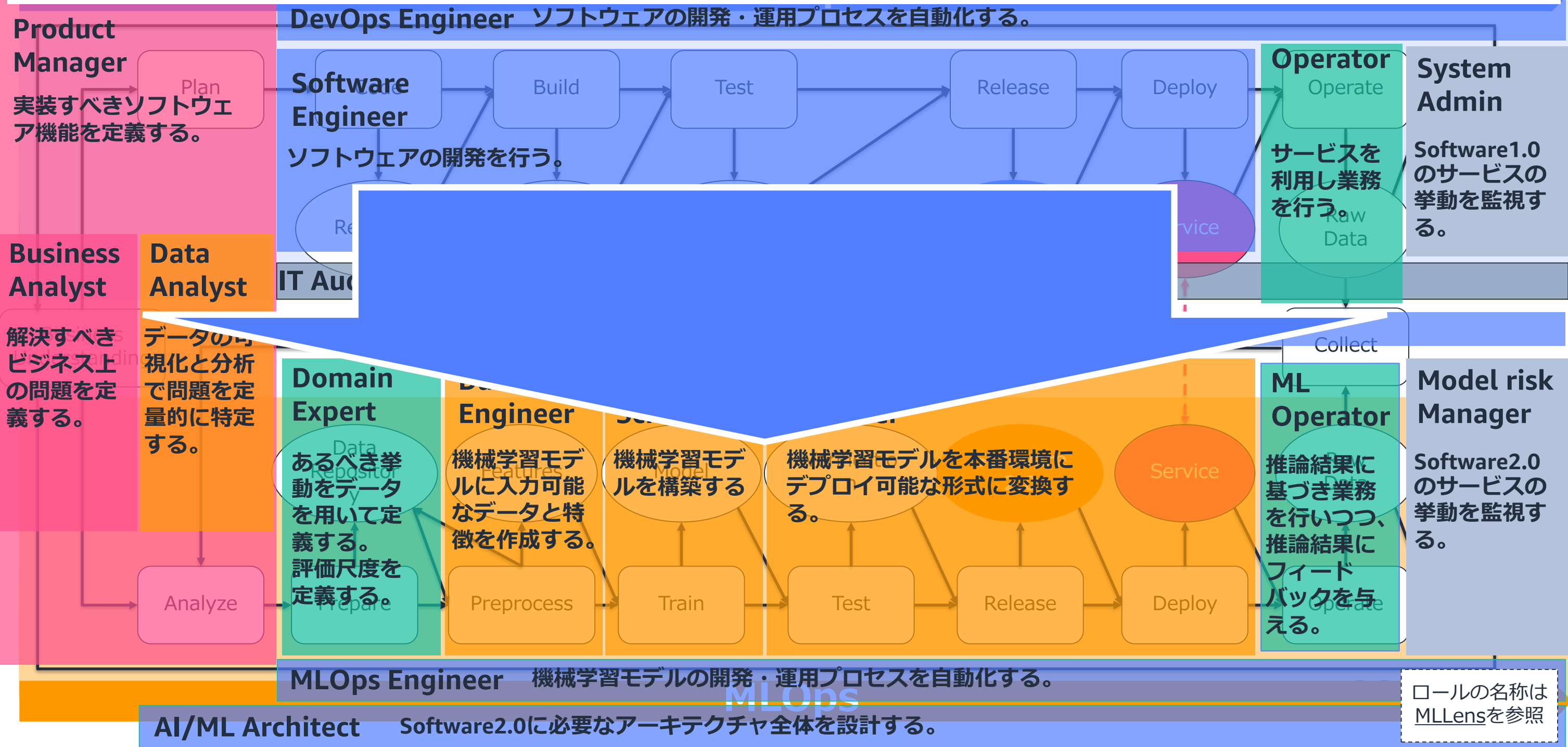
機械学習モデルの開発・運用プロセスを自動化する。

AI/ML Architect

Software2.0に必要なアーキテクチャ全体を設計する。

ロールの名称はMLLensを参照

DevOps側のチームがMLOps側の足りないロールをカバーしたり、マネージドサービスを活用する必要がある。



実装だけでなくコミュニケーションが重要

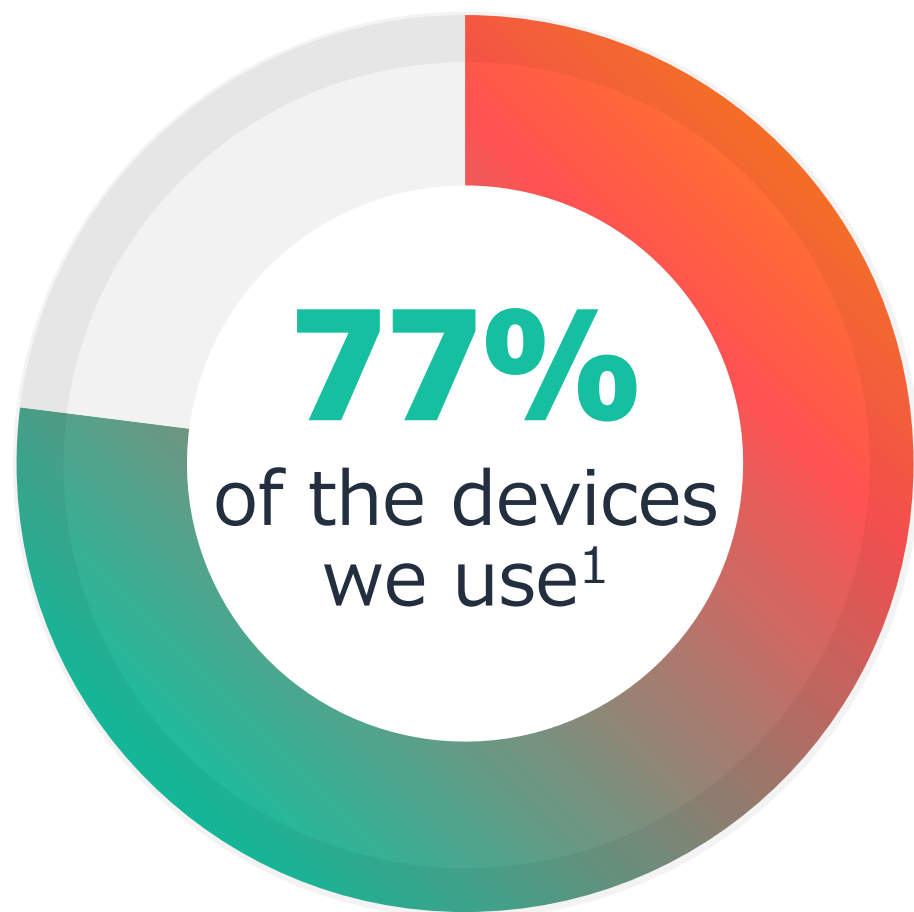
- Pythonで機械学習の基本的な実装を行えるようになる。
- 開発に関わる他のステークホルダーに、Jupyter Notebookを提示しながら必要な確認をとれるようになる。
- データサイエンティストに対し適切な声掛けができるようになる。

本日解説すること

1. 機械学習モデルの開発プロセスと組織
2. 機械学習を使うメリット・デメリット
3. 開発プロセスを解説する流れ

**苦勞して
機械学習を使ったほうがいいんですか？**

顧客にとってAI/MLの機能があることは当たり前になりつつある



ライドシェア (例: Uber, Lyft)

待ち時間、需要予測、価格設定



オンラインショッピング

検索結果、商品レコメンド



金融機関

書類（小切手等）の認識
クレジット・デビットカードの不正取引検知



2023年の世界

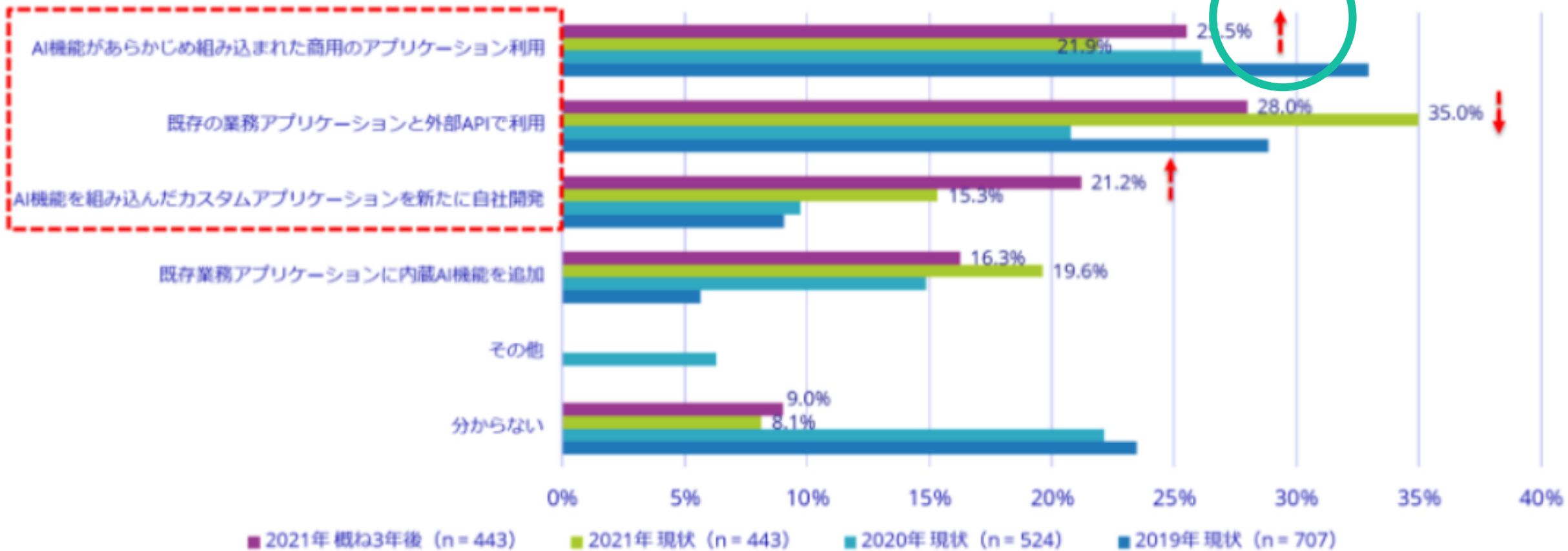
80億個以上のAI/ML搭載デジタルアシスタント

[101 Artificial Intelligence Statistics \[Updated for 2022\]](#)

[DIGITAL VOICE ASSISTANTS IN USE TO TRIPLE TO 8 BILLION BY 2023, DRIVEN BY SMART HOME DEVICES](#)

AI機能が組み込まれたアプリケーションの利用意欲は増加傾向

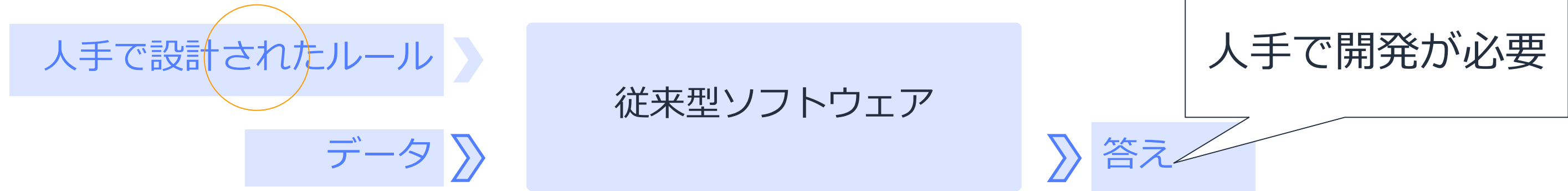
Q. AIシステムを利用する場合、どのような形態をとりますか？
また、概ね3年後にどのような形態をとる予定ですか？



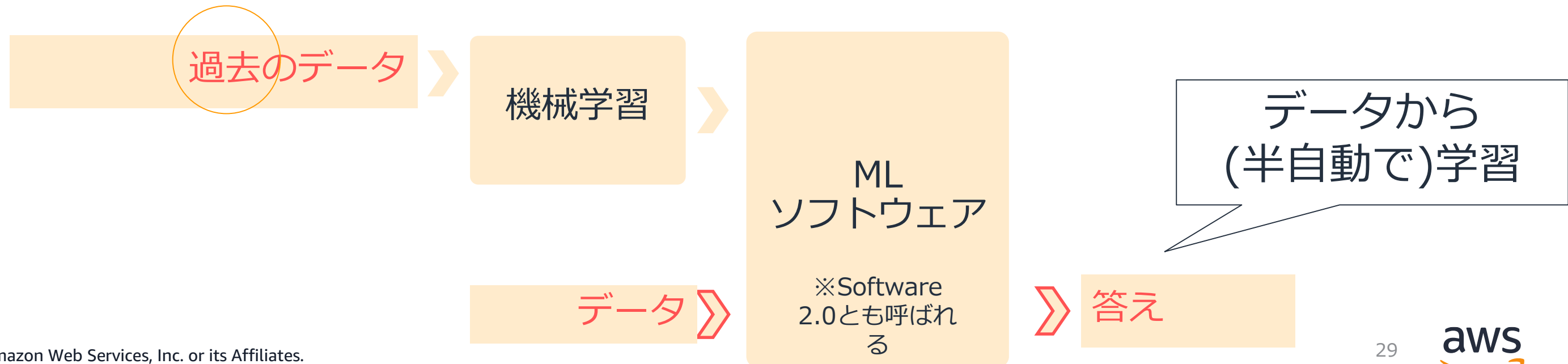
IDC 「2021年 国内AIシステム/アナリティクス市場 企業ユーザー調査」

プロダクトで機械学習を使うメリットは、 顧客自身のデータから**半自動**で顧客体験を改善できること。

従来のソフトウェア



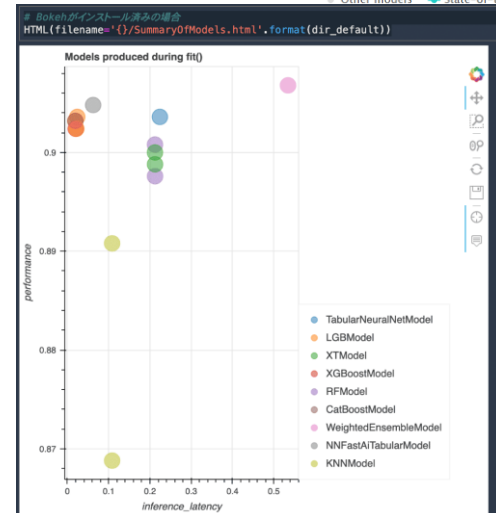
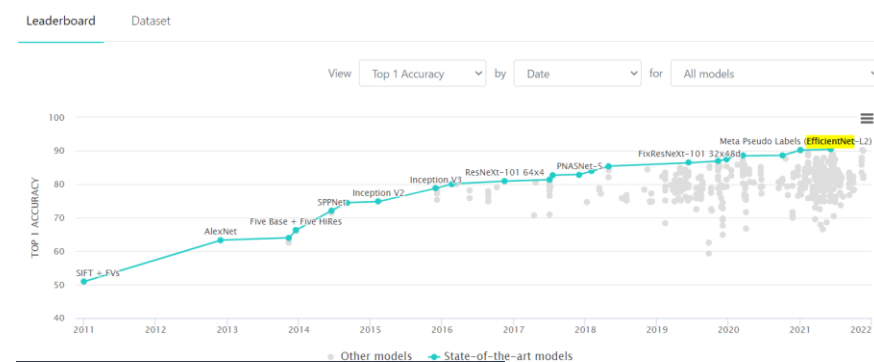
機械学習モデル



AutoMLの進化が「半自動化」を加速している。

機械学習モデルを自動で構築するAutoMLの技術が進化し、データサイエンティストが構築せずとも高い精度のモデルが構築できるようになってきている。

Image Classification on ImageNet

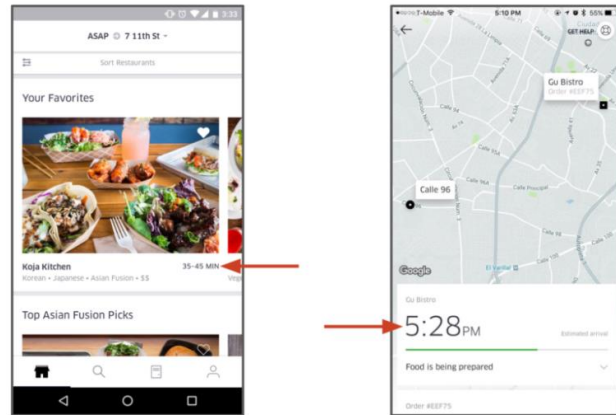


画像分類で現在上位の分類性能を記録しているモデルは、**AutoML技術により構築されたモデル**([Efficient Net](#)など)。自然言語処理の分野では、大規模なテキストで事前学習することで少ないデータしかないタスクでも高い精度が記録できることが確認されている([BERT](#)など)。

OSSのAutoMLライブラリも開発され、手軽に利用できる。 AutoGluonではOtto Group[2015年]とBNP Paribas[2016年]のKaggleコンペティションで、**参加したデータサイエンティストの99%よりも高い精度を記録**([論文](#))。

([開発者インタビューの記事があるので興味がある方はぜひ](#))

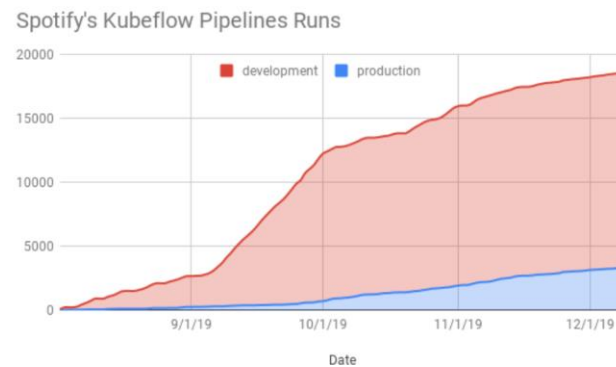
ソフトウェアが使われるほどデータが増え、データが増えるほどモデルが高精度になり、より使われるループを構築できる。



画像引用: [Meet Michelangelo: Uber's Machine Learning Platform](#)

Uber Eats

機械学習を用いて配達時間を予測。配達時間が正確になるほど顧客体験が改善され、顧客体験が改善されるほど利用者が増えてよりデータが集まり正確になる。

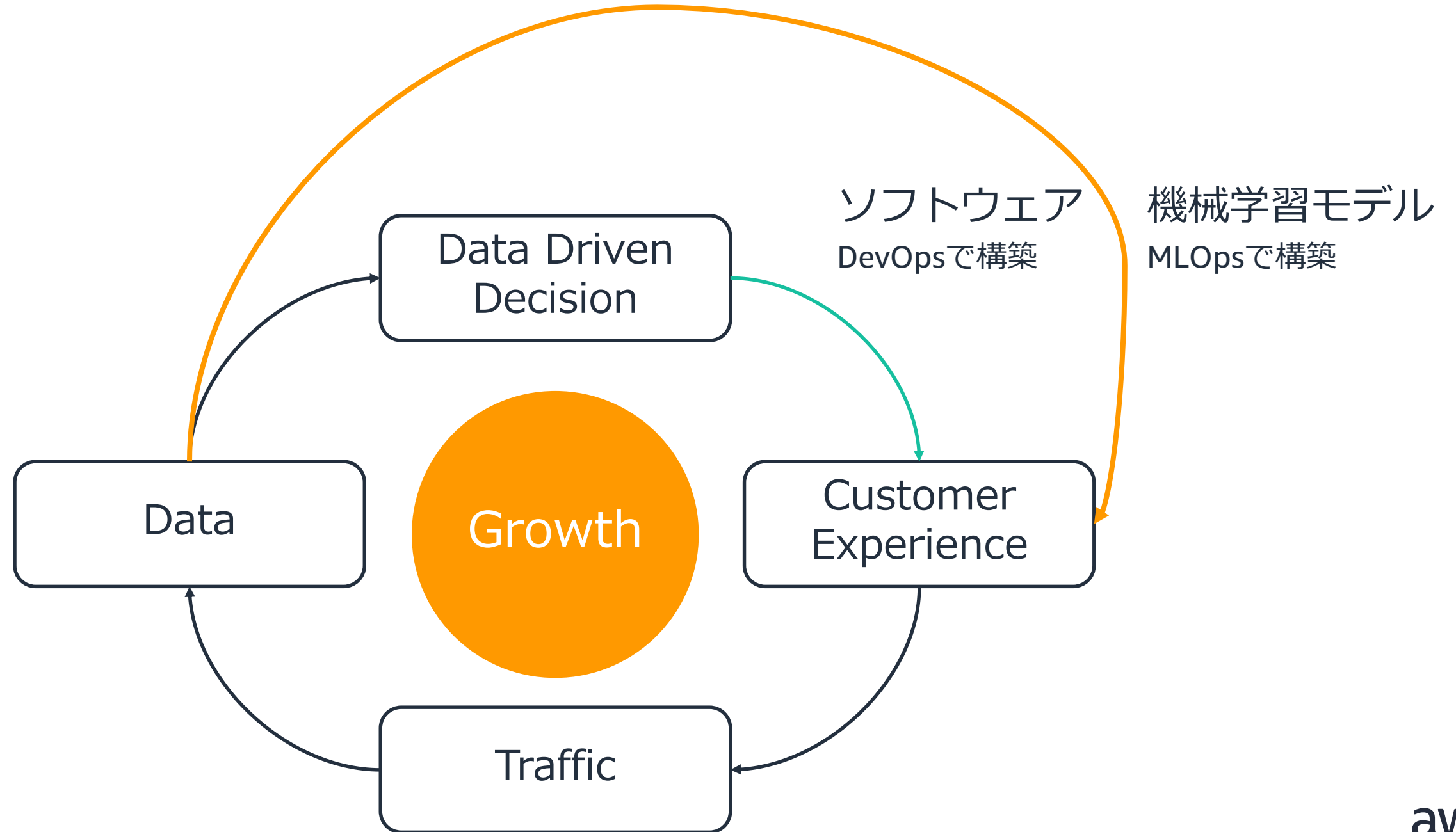


画像引用: [The Winding Road to Better Machine Learning Infrastructure Through Tensorflow Extended and Kubeflow](#)

Spotify

機械学習を用いて新しい音楽と出会うDiscover Weeklyなどの機能を実装。本番の機械学習ワークロードの約5倍の実験を行うことで新しい機能の開発を行っている(通常のソフトウェア開発では困難な開発スピード)。

データの蓄積があれば従来型のソフトウェアより高速かつ正確に顧客体験を改善できる。



もちろん、機械学習にデメリットはある。

- **一定量の教師データが必要とされる**
(多くの場合)タグ付けされた同一形式のデータを一定量必要とする。
- **データに偏りがあれば偏りのあるモデルが生成される**
画像認識システムが黒人の写真に「ゴリラ」というタグを付けた。
白人より黒人、男性より女性のジェンダーの分類を誤りやすい、など。
- **必ずしも説明性を持たないモデルもある**
説明が必要: ローン申請に使用しており、拒絶された顧客に対する説明。
許容される: 最終判断が人間によって行われるとき。

実装するうえでの難しさもある。

動画を通じ解決方法をお伝えしていきます。

1. **Wide**

データの取得、前処理、学習など幅広い処理を繋ぎ合わせ実装するため、統合的なプロセス管理やバージョン管理が必要。

2. **Large**

開発環境に要求されるリソースが大きいため新しい環境構成が必要。

3. **Probabilistic**

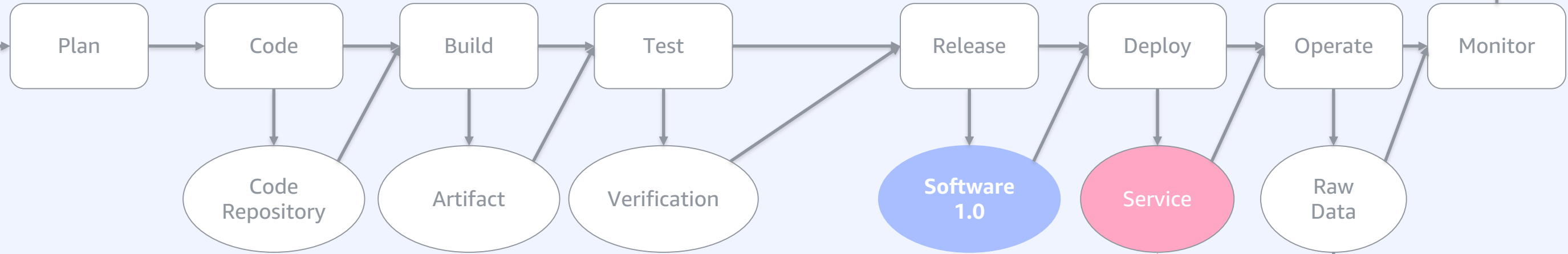
確率的な挙動をするため新しい品質ガイドラインに基づくテストが必要。

本日解説すること

1. 機械学習モデルの開発プロセスと組織
2. 機械学習を使うメリット・デメリット
- 3. 開発プロセスを解説する流れ**

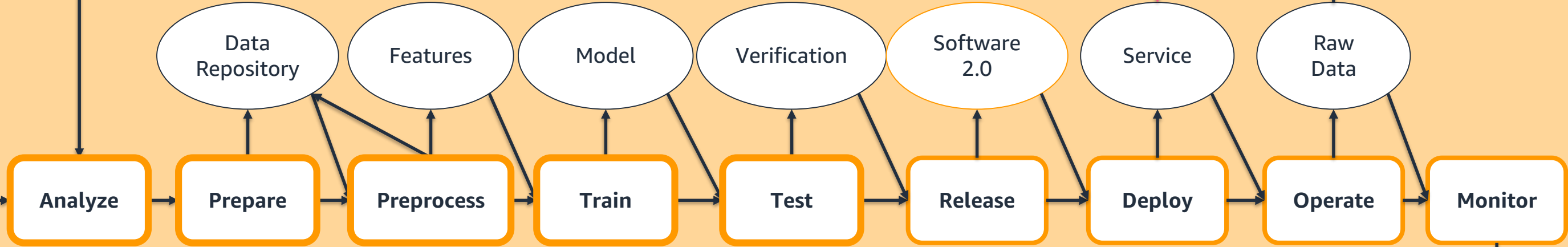
機械学習の開発プロセス(MLOps)を解説していきます。

DevOps



Business Understanding

Data Warehouse



MLOps

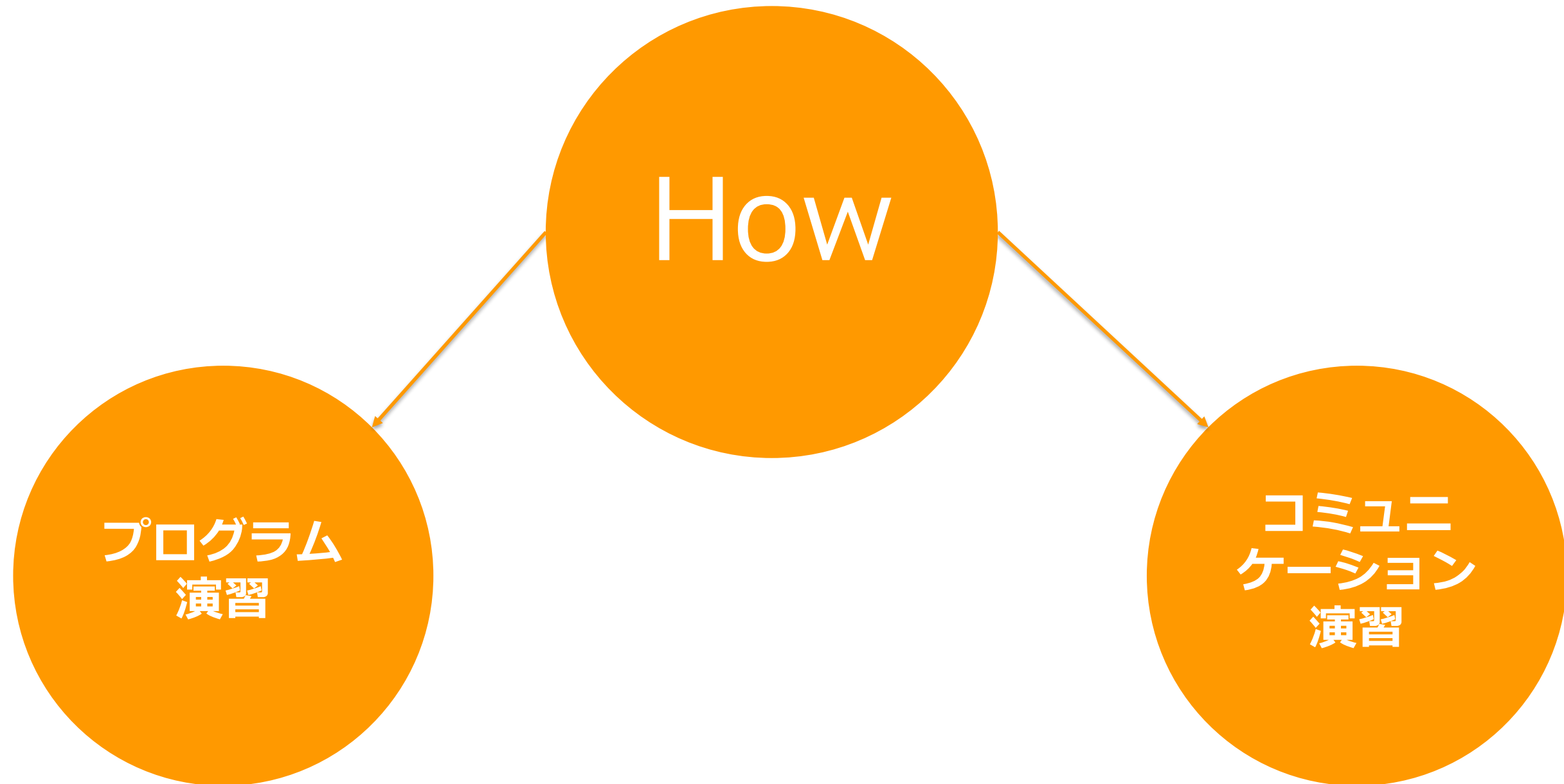
ML(開発)とOps(運用)のうち、開発の6プロセスをまず解説していきます。

英語名	日本語名	なにをするのか
Business Understanding	ビジネス理解	ビジネス上発生している問題を機械学習モデルの予測がどのように解決するのか、関係者へのヒアリングを通じ理解しBusiness goalを設定する。
Analyze	データ分析	Business goalを達成するために現状手に入るデータが必要十分であるか、データ分析を行いAnalysis Reportを作成する。
Prepare	データ準備	Analysis Reportをステークホルダーに共有し、必要なデータや不足しているデータの収集や作成を行いDatasetを作成する。
Preprocess	前処理	Datasetを機械学習モデルが認識しやすい形に変換し、Preprocessed Datasetを作成する。必要に応じデータ同士を組み合わせる新しいデータを作るFeature Engineeringを行う。
Train	学習	Preprocessed Datasetを用い、Modelを学習する。
Test	テスト	Modelが意図した通りに構築されているか、インプット、プロセス、アウトプットの観点から評価しTest Reportを作成する。Test Reportには、Business goalを達成できているかの評価を含める。

各プロセスの解説は、なにをするのか、なぜやるのか、どうやるのかの3つから構成されます。



どうやるのかは、プログラミングだけでなくコミュニケーションの演習も行います。



コミュニケーション演習では、他のロールの人に作業内容の確認をするための質問を作る演習を行います。



本動画シリーズで概要、別シリーズで実装の詳細を解説します

Light Part (本動画シリーズ)

Introduction

Environment Setup

1. Business Understanding
2. Analyze
3. Prepare
4. Preprocess
5. Train
6. Test
7. Release
8. Deploy
9. Operate
10. Monitor

機械学習モデルの
開発プロセスの概
要を解説

Dark Part

Environment Setup

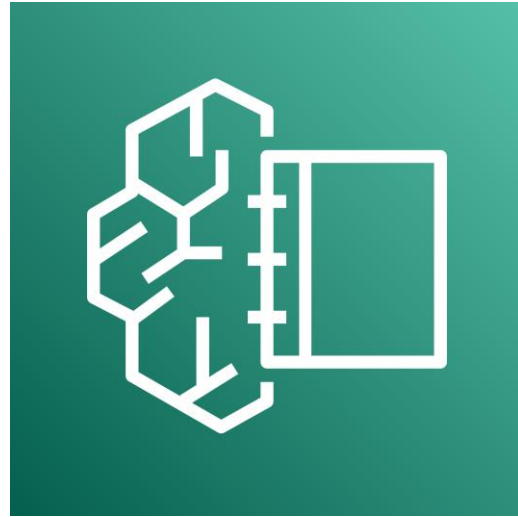
- Business Understanding
- Analyze
- Prepare
- Preprocess
- Train
- Test
- Release
- Deploy
- Operate
- Monitor

個別のプロセスに
ついて実装方法の
詳細を解説

Light Partの学習に費用は一切かかりません。

Light Part

SageMaker Studio Lab



機械学習を学び、体験し、検証する

Learning

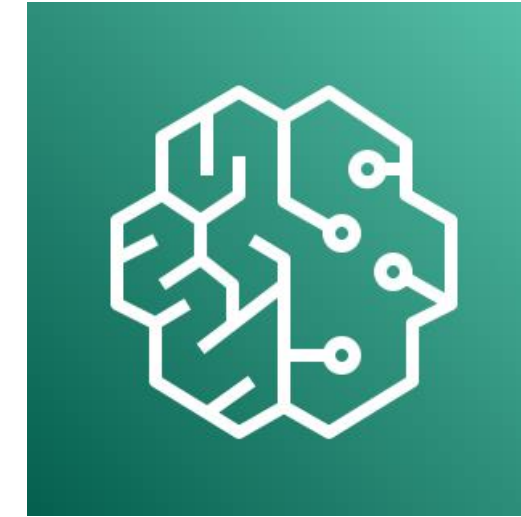
Experimenting

Prototyping

無料+AWSアカウント不要

Dark Part

SageMaker



機械学習モデルを構築・運用・スケールする

Building

Deploying

Scaling

本格的な実装方法はDarkで！

エクスポートが可能

Dark PartではAWSを利用した実装方法を解説します。

- 例: Trainのフェーズで、大規模な機械学習モデルを学習するのにAWSを利用する方法など。



資料はすべてオープンな場所で公開しています

動画

資料

実装

YouTube

Speaker Deck

GitHub

<https://github.com/aws-samples/aws-ml-enablement-handson> をチェック！

フィードバックをお待ちしています！

<https://github.com/aws-samples/aws-ml-enablement-handson>

aws-samples / aws-ml-enablement-handson Public

<> Code Issues Pull requests Discussions Actions Projects Wiki Security Insights Settings

main 1 branch 0 tags Go to file Add file Code

icoxfog417 Update README.md 797ac75 5 days ago 3 commits

folder	.github/ISSUE_TEMPLATE	Add hands-on contents	5 days ago
folder	docs	Add hands-on contents	5 days ago
folder	model	Update README.md	5 days ago
folder	notebooks	Add hands-on contents	5 days ago
file	.gitignore	Add hands-on contents	5 days ago

資料集・お問合せ・Special Thanks

AWSの日本語資料の場所: 「AWS 資料」で検索



お問合せ

- 技術的なお問合せ
- 料金のお問合せ
- 個別相談会のお申込み

AWSのハンズオン資料の場所: 「AWS ハンズオン」で検索



Special Thanks

- 音楽素材: PANICPUMPKIN様

SEE YOU NEXT VIDEO!!

Light Part (本動画シリーズ)

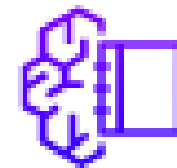
Introduction

Environment Setup

1. Business Understanding
2. Analyze
3. Prepare
4. Preprocess
5. Train
6. Test
7. Release
8. Deploy
9. Operate
10. Monitor

プログラミング演習

- 機械学習モデルを構築するための環境を作る方法を学びます。



amazon

SageMaker Studio Lab

コミュニケーション演習

- 環境を構築するとき他のロールに確認すべき事項を学びます。