



ML Enablement Workshop #3 : 開始編

顧客体験の改善を開始する

久保 隆宏

Developer Relation
Machine Learning

ML Enablement Workshopのプロセス

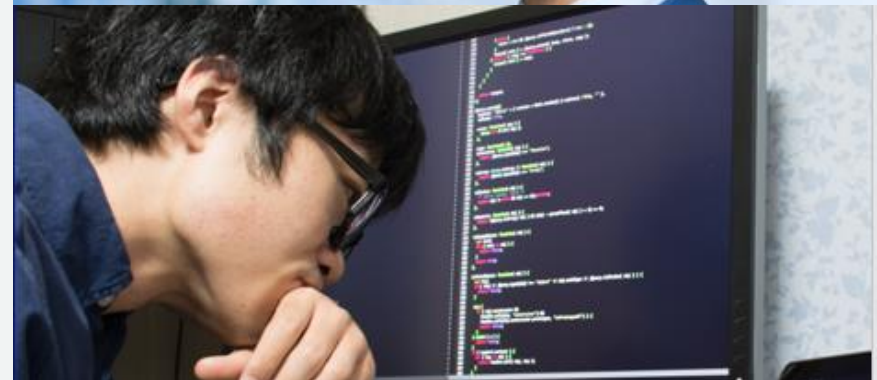
理解: ベストプラクティスの理解

機械学習の「勝ちパターン」を理解し、活用どころを見極める。



応用: 顧客体験改善への応用

勝ちパターンが成立するプロセスを可視化し解決すべき課題を特定する。

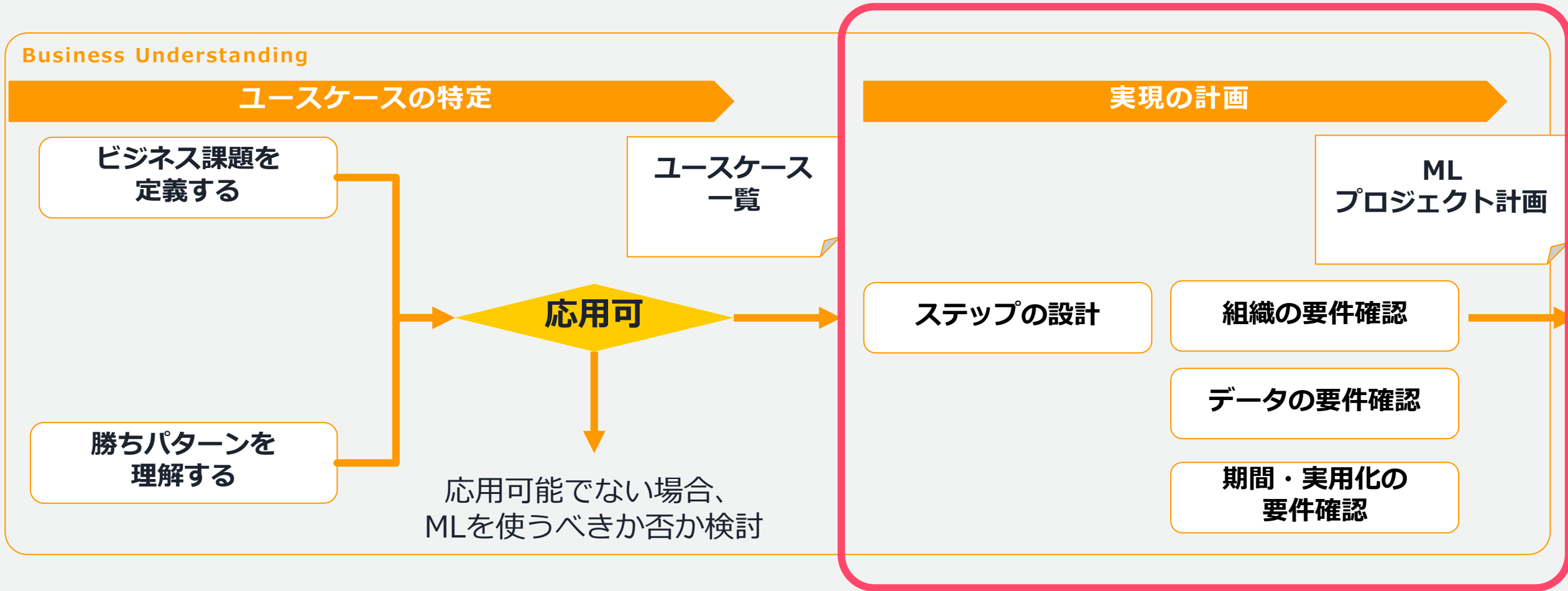


開始: 顧客体験の改善を開始する

機械学習による勝ちパターン成立まで、段階的に顧客体験の改善をする。



ビジネス理解: Business Understanding



「開始」プロセスの対象者とゴール

誰のためのセッション?

機械学習の活用に取り組むプロダクト開発チーム
プロダクトマネージャー、
開発者、
データサイエンティスト
から構成される。

何ができるようになる?

データから半自動で顧客体験を改善する機械学習の活用をゴールとし、
段階的にステップアップする計画を立てられるようになる。

何ができるようにはならない?

機械学習アルゴリズムの実装。

疑問：機械学習の実装がはじまらないのはなぜ？

回答：不確実な成果と確実なコストの比較結果

詳細は本編で！

1. 機械学習プロジェクトの不確実性

2. ステップの設計

3. さらに先へ進むための要件確認

機械学習の導入効果は不確実だが、開発する人員とマシンのコストは常に確実

特に初めて機械学習を導入する場合、4つの不確実性が存在する。

1. 機械学習のモデルが開発できるか
2. 機械学習のモデルの精度が予想通り達成できるか
3. 顧客の問題が解決されるか
4. ビジネスの問題が解決されるか

機械学習の導入効果は不確実だが、開発する人員とマシンのコストは常に確実

特に初めて機械学習を導入する場合、4つの不確実性が存在する。

1. 機械学習のモデルが開発できるか
2. 機械学習のモデルの精度が予想通り達成できるか
3. 顧客の問題が解決されるか
4. ビジネスの問題が解決されるか

効果を検証する前にモデルの開発コストを支払う必要がある

回答: 不確実な成果と確実なコストの比較結果

モデルを開発しないと効果は計測できないのか？

機械学習モデルで実現したい機能を人手・従来型ソフトウェアで代替することで開発した場合の効果を実測できる。

人手・従来型ソフトウェアで代替

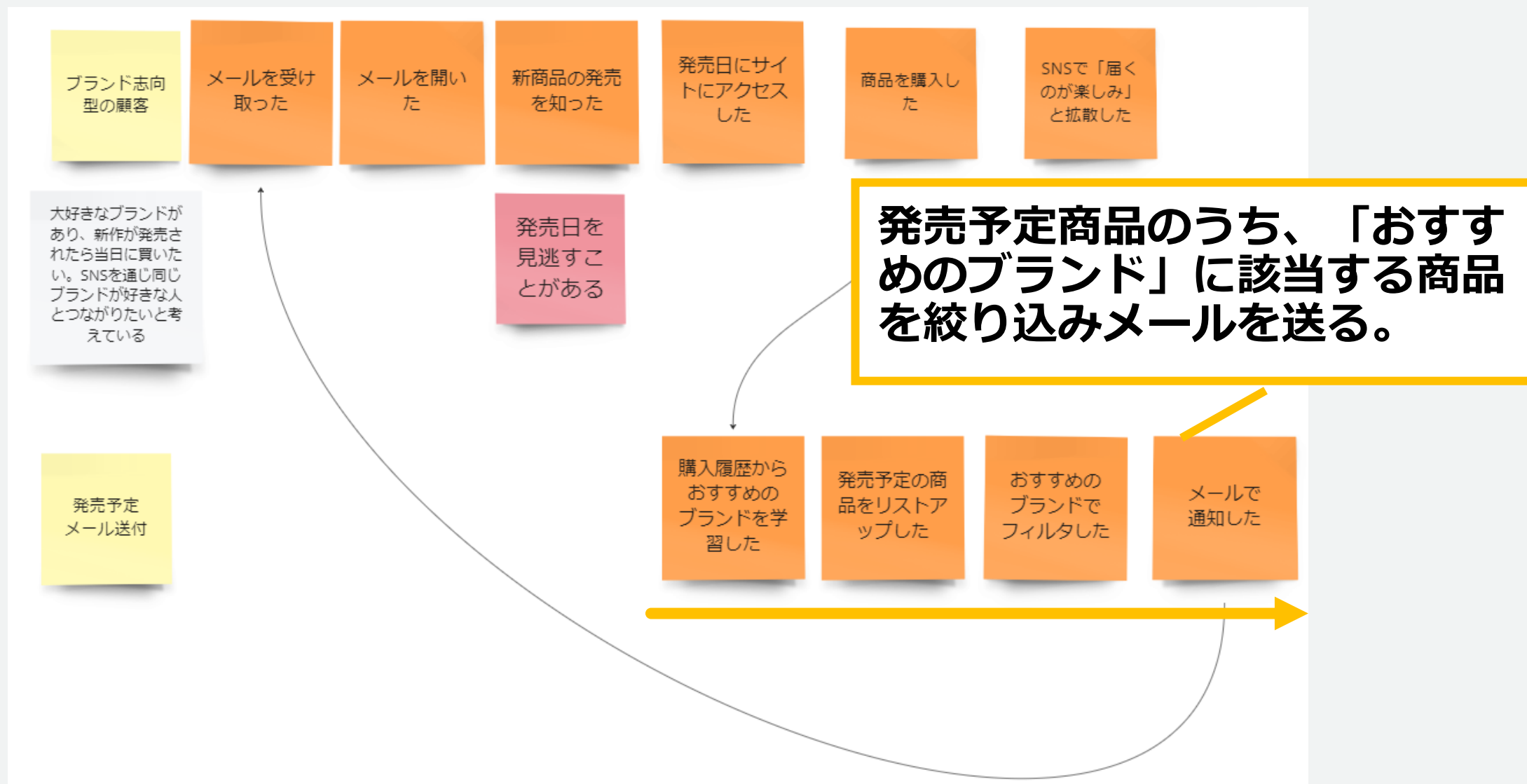
- 3. 顧客の問題が解決されるか
- 4. ビジネスの問題が解決されるか

1. 機械学習プロジェクトの不確実性
- 2. ステップの設計**
3. さらに先へ進むための要件確認

例：ファッションサイトでパーソナライゼーションを導入する



パーソナライゼーションの導入案



どう段階的に効果を確認し機械学習へ向かうか

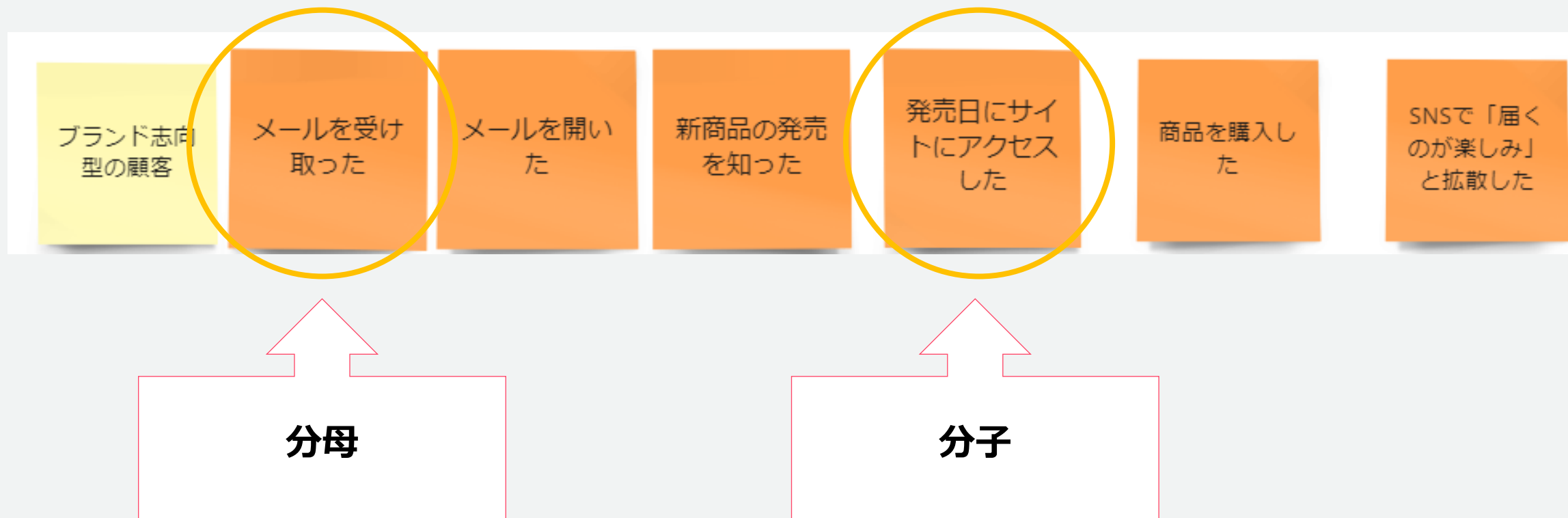
1st Step : データに基づくマーケティングメールの送付からはじめる

データに基づく メール送付の実施

マーケティングチームが、ブランドごとに購入頻度が高い顧客を絞り込み新作発売前にメールを送る。

顧客体験の計測: メールからのアクセス数 / メール送付数

可視化したイベントから顧客体験を計測する



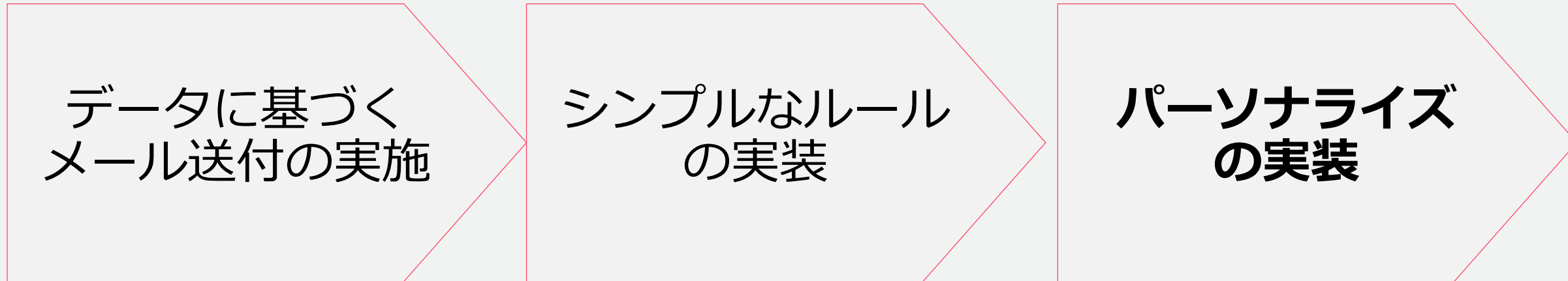
2nd Step: シンプルなルールを実装し規模を拡大



開発チームが、新作発売前のメール送付を効率的に行えるよう、クエリの実行によるメール送付先リスト作成を自動化。

顧客体験: メールからのアクセス数 / メール送付数

3rd Step: パーソナライゼーションの実装



データサイエンティストが、メールからのアクセス数、購入履歴からより精度高く送付先を特定するパーソナライゼーションを作成。

顧客体験: **メールからのアクセス数** / メール送付数

どう段階的に効果を確認し機械学習へ向かうか

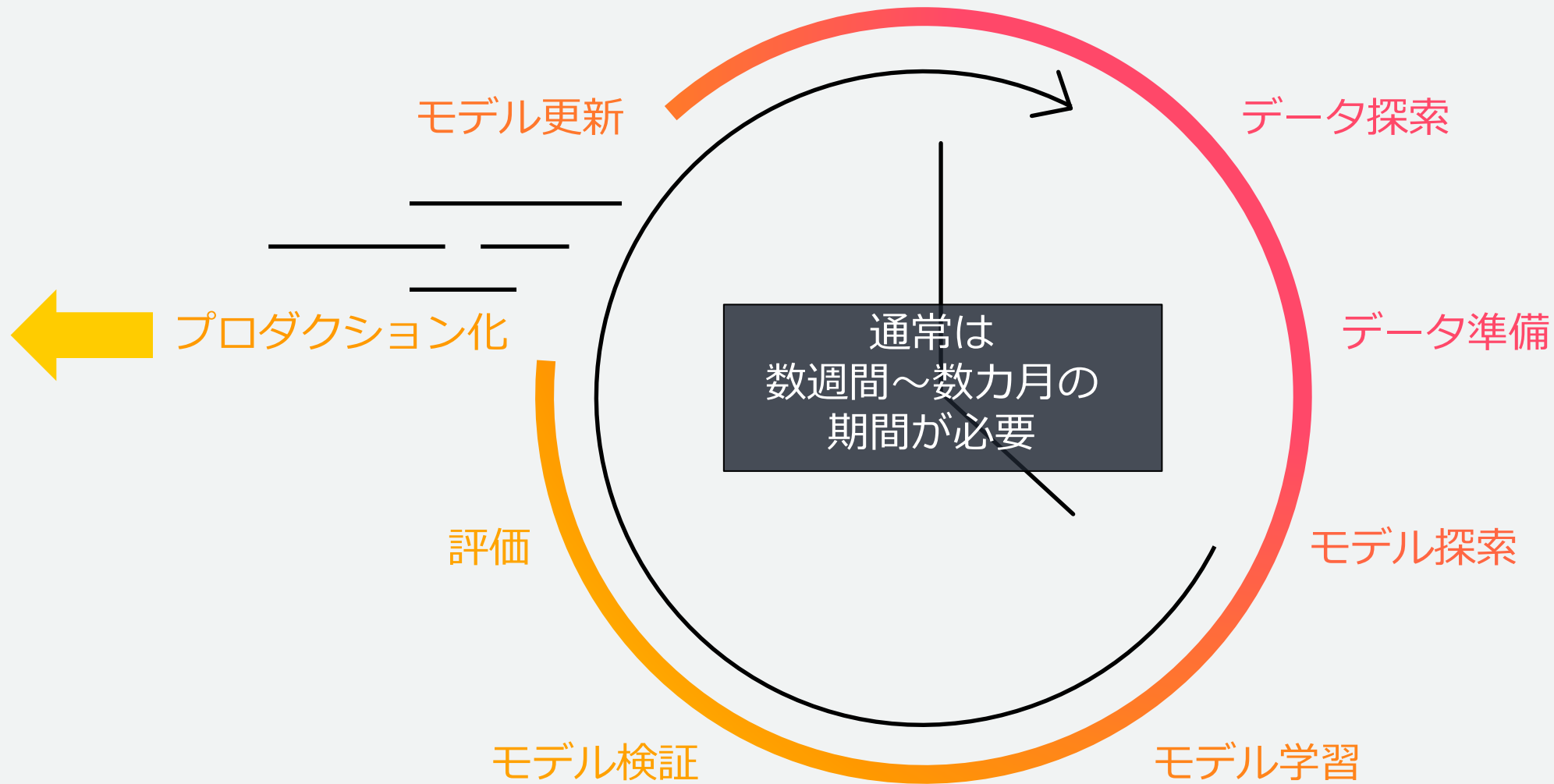
データに基づく
メール送付の実施

シンプルなルール
の実装

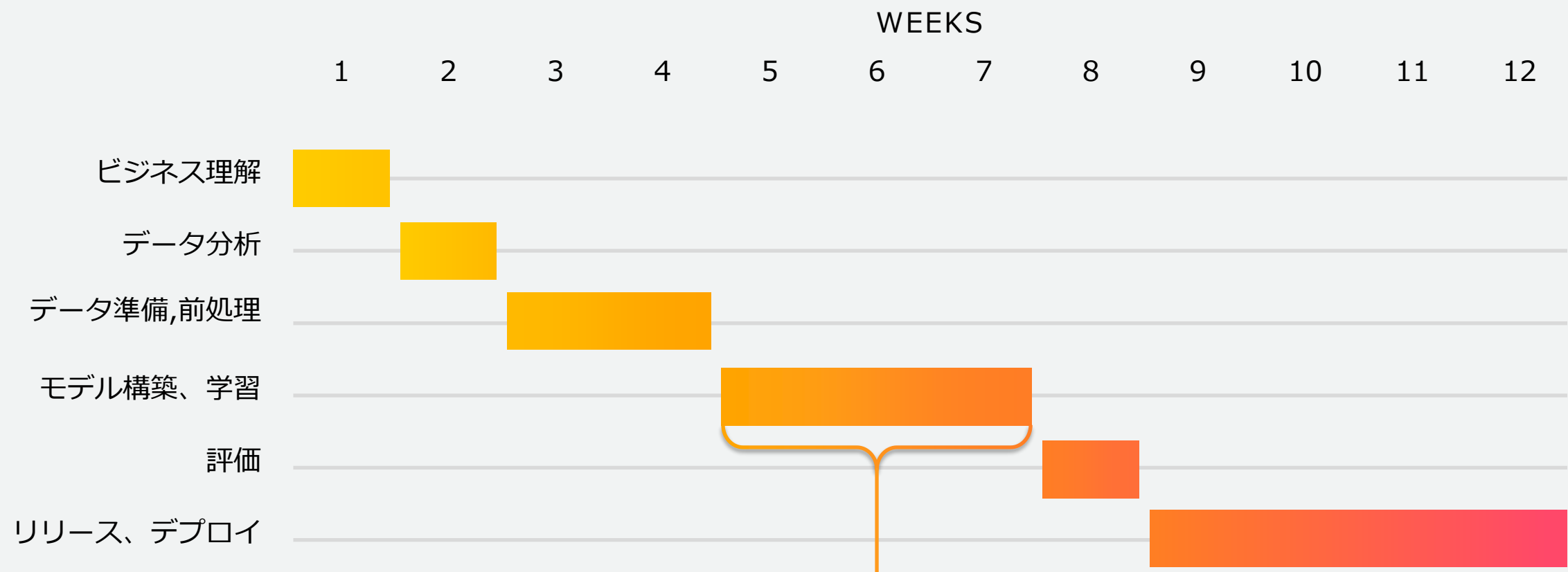
パーソナライズ
の実装

「業績志向」で、段階的にビジネス成果を確認しチームと投資を拡大する。

機械学習の段階へ進む際の注意点: プロジェクト期間計画



機械学習の段階へ進む際の注意点: プロジェクト期間計画



よくある誤り :
機械学習モデル関連の期間のみを見積る

プロジェクトの遅延要因

前頁はあくまでベストシナリオであり下記の要因により追加の期間が必要となる



前例のないタスク

深い科学的な知見が必要



探索時にデータの不備や
ラベル不足が発覚

解消に数カ月を
要することも



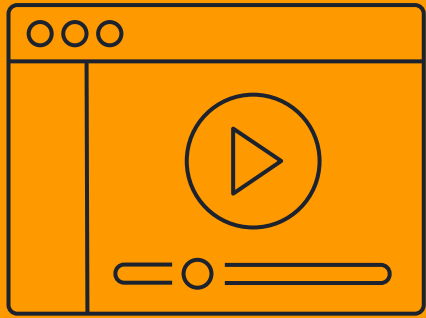
モデルの性能が
ビジネス要件を満たせない

追加実験の
繰り返しが必要

本番導入にあたっての疑問を早めに意識する

- › 推論にかけられる時間、計算コストは？
- › データはどのくらいの頻度で変わる？
- › プロダクション化のために必要な変化は？
- › モデルの性能はビジネスニーズを満たしている？

ケーススタディ：データ関連の遅延



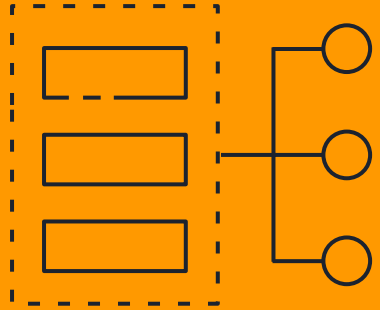
ゴール:

ある製造業が製品の**画像をデータベース化**しており、
画像に基づく**不良品の検出**を
実現したい

課題:

画像データは製造ライン上に配置された
カメラにより、1カ月間自動的に撮影さ
れていたが、**アノテーションが付与され
ていない**

ケーススタディ：データ関連の遅延



結論：

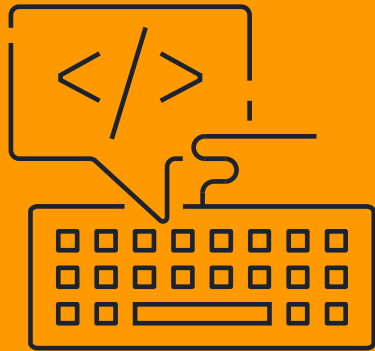
MLモデルの学習のためには画像に対するラベリングが必要

ラベリングのためには、画像を正確に見極められる製造プロセスのエキスパートが必要

教訓：

- › MLプロジェクトの多くはラベル付きのデータが必要だが、容易にラベリングできないこともある
- › **データとともにラベルを確実に収集することにより、プロセスを簡略化できる**

ケーススタディ：実用化の要件を満たすための遅延



実験用コード

探索的

反復的

テスト、エラー対策少



プロダクション用コード

要件の確定

バージョン管理

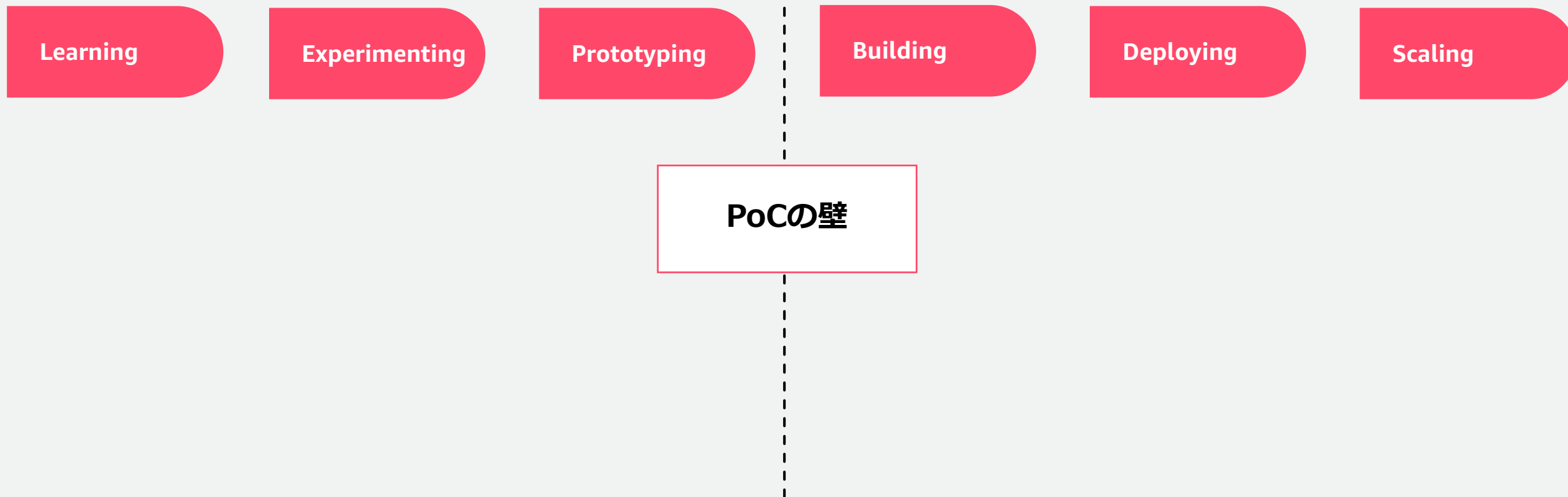
実用レベルの信頼性

1. 機械学習プロジェクトの不確実性
2. ステップの設計
- 3. さらに先へ進むための要件確認**

組織におけるMLジャーニー

機械学習を学び、検証し、試作する

機械学習モデルを構築・運用・スケールする



MLジャーニーの推進に必要な要件

AI活用へのシフトを成功させるために必要な組織としての要件



堅牢なAI戦略



データ戦略



組織内のコラボレーション

これらを推進するためのリーダーシップ

1. 堅牢なAI戦略

2. データ戦略

3. 組織内のコラボレーション

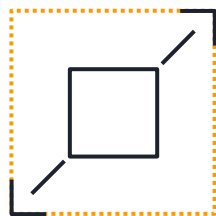
4. 推進のためのリーダーシップ

AI戦略における主な検討事項



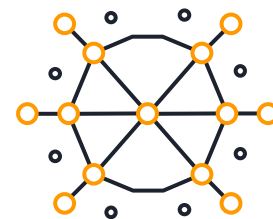
クラウドの力を活用

クラウドの活用により、PoCの開発を加速させる。



PoCを越えるスケーラビリティ

PoCの先を見据えたイメージを作る。新たなプロジェクトを高スケールで実用化する計画を立てる。



組織文化の変革

貴組織のニーズに合わせたMLチームをどのように作れるかを考える。

戦略: クラウドの力を活用



MLを簡単に

ワンクリックで学習・デプロイ、ビルトインアルゴリズム、自動チューニング



スケーラブル

強力なGPUへのオンデマンドアクセス、オートスケーリング、分散学習



コストパフォーマンス

トータルコスト(TCO)の削減、利用分だけ支払う



セキュア

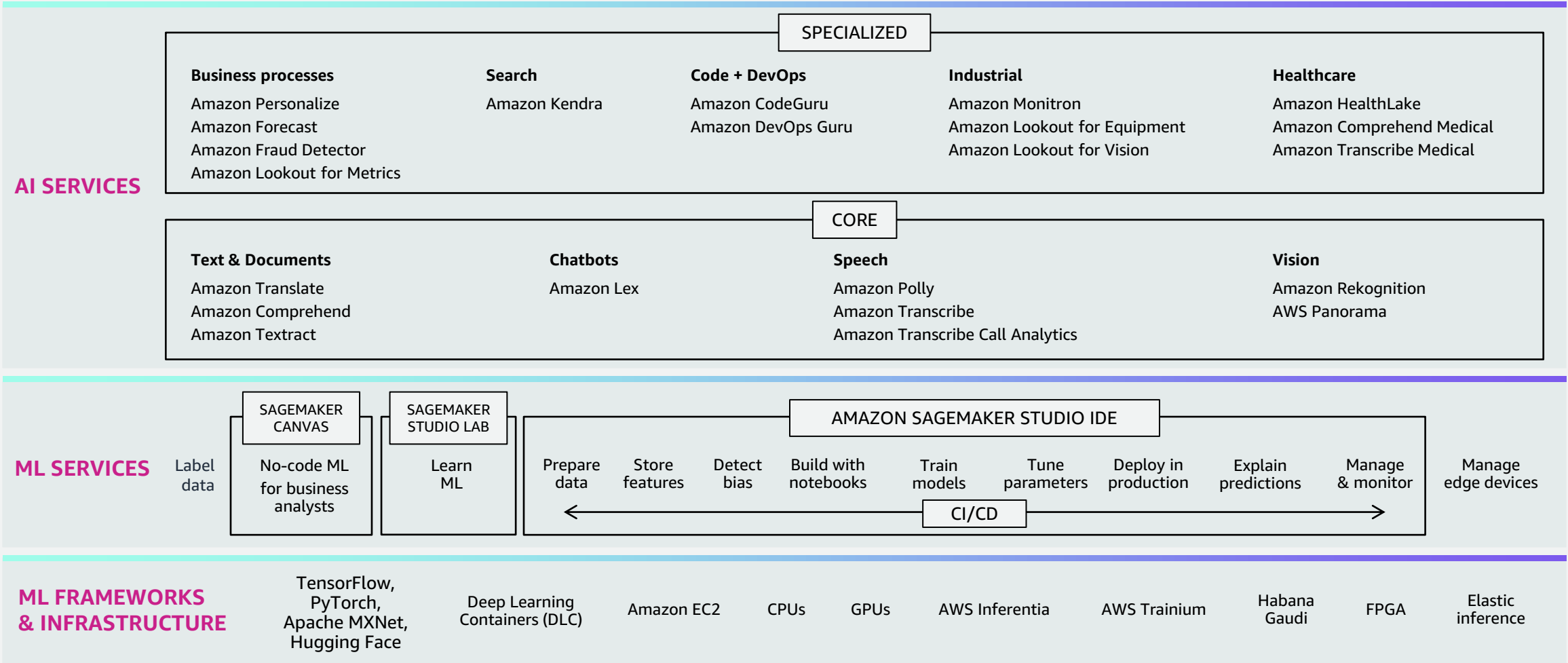
完全なセキュリティ機能: ネットワーク隔離、アクセス制御、暗号化、コンプライアンス



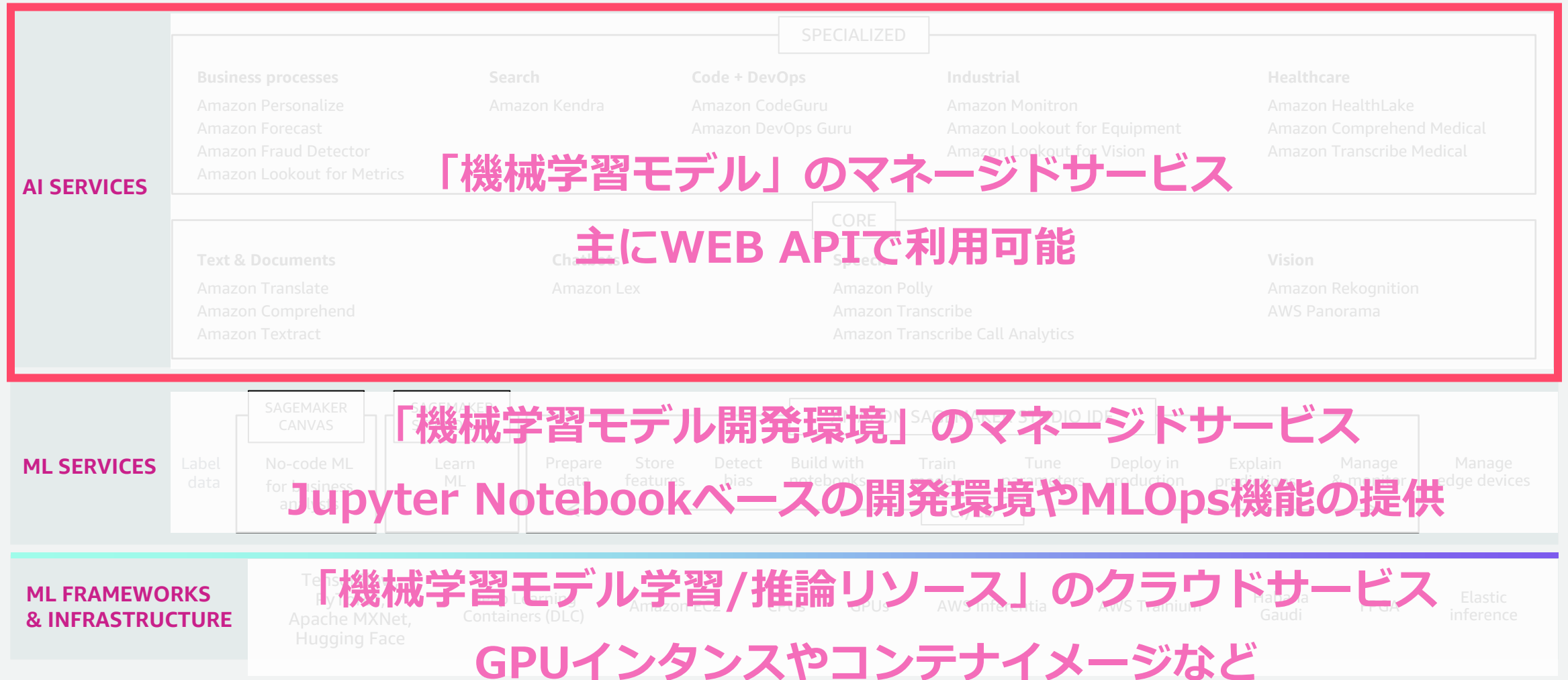
DevOps対応

フルマネージドMLOps、Kubernetesでの稼働オプション

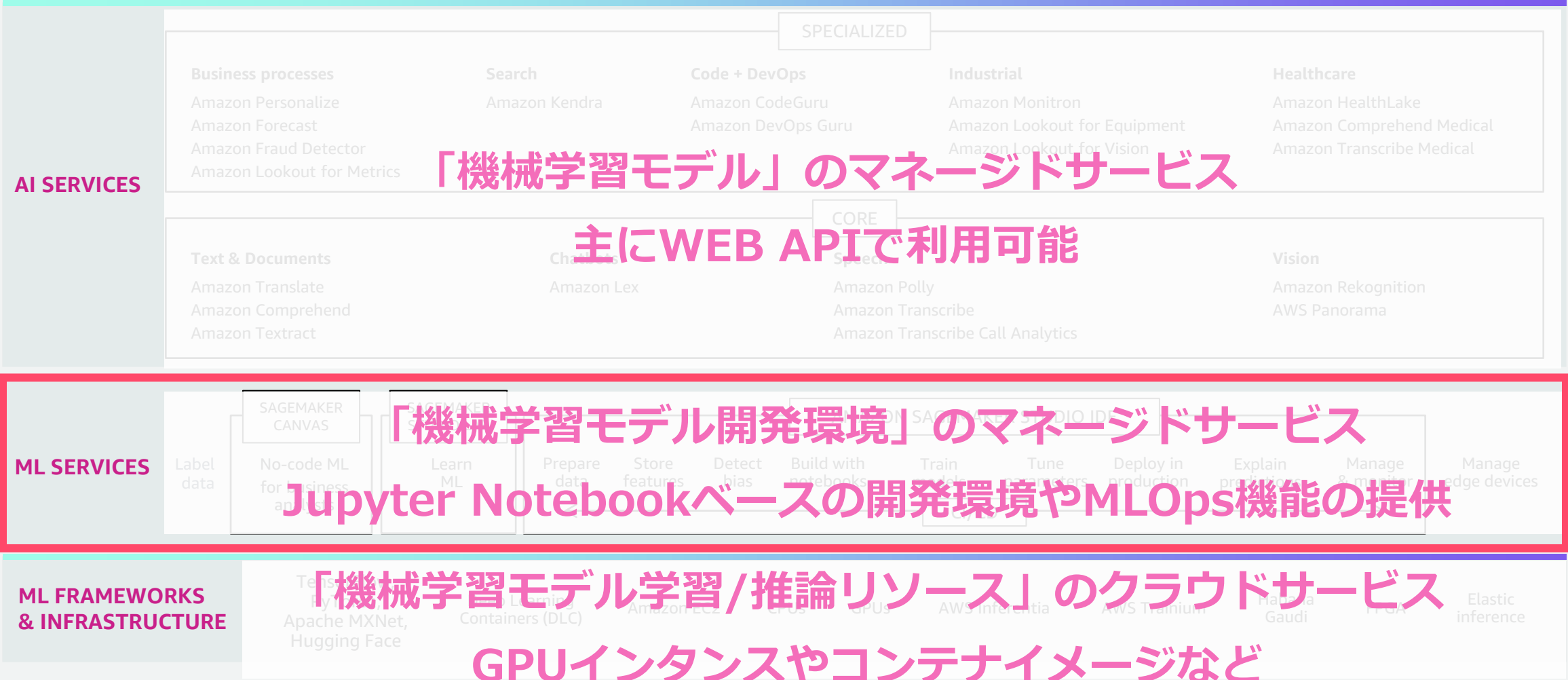
戦略2: クラウドの力を活用



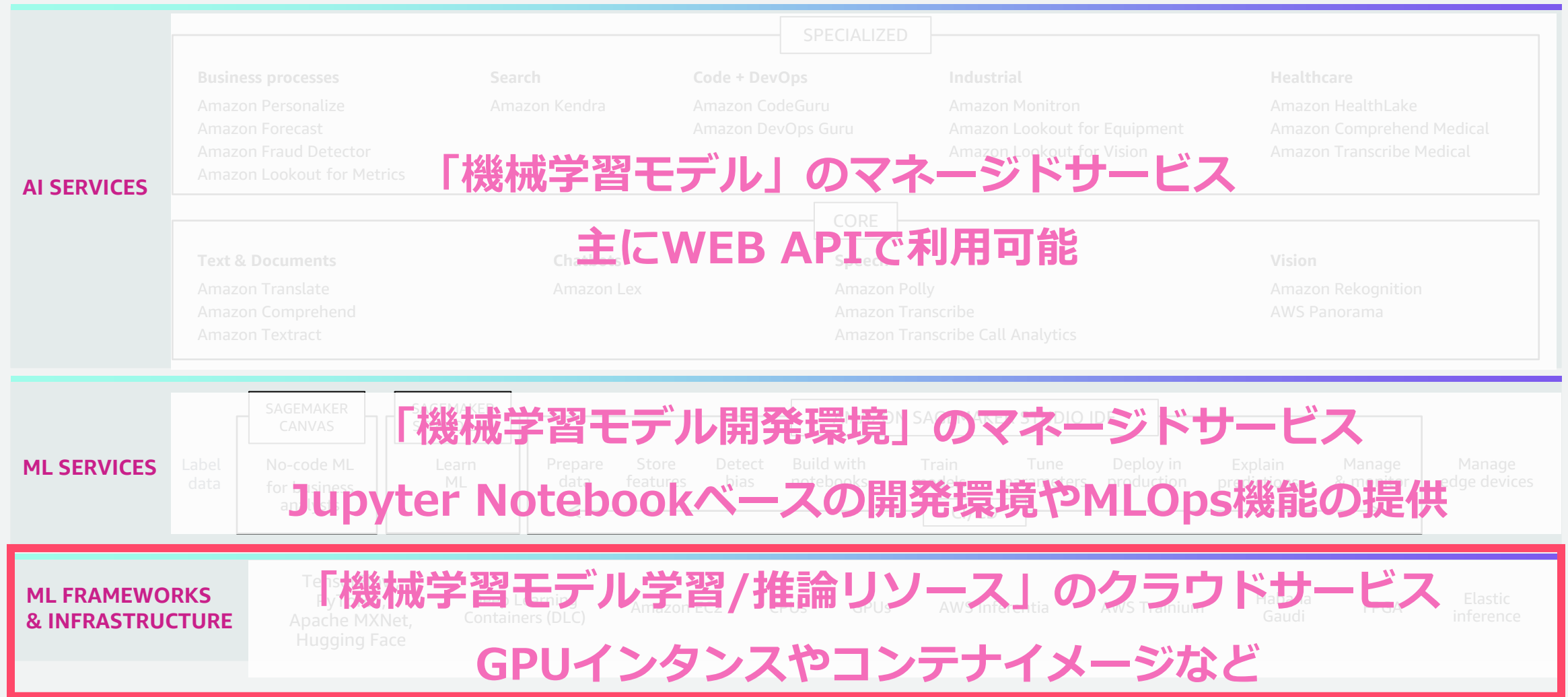
戦略2: クラウドの力を活用



戦略2: クラウドの力を活用



戦略2: クラウドの力を活用



戦略2: クラウドの力を活用

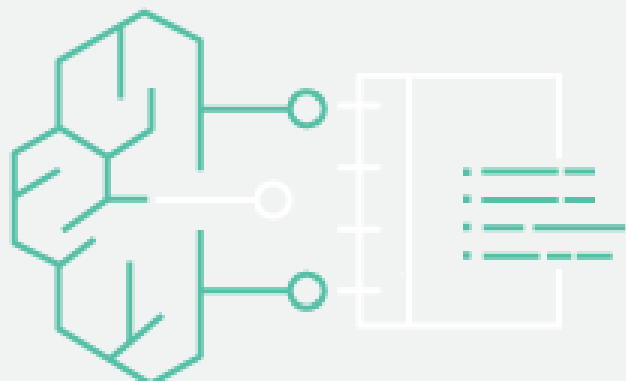
機械学習を学び、検証し、試作する

Learning

Experimenting

Prototyping

SageMaker Studio Lab



無料・メールアドレスのみで利用可

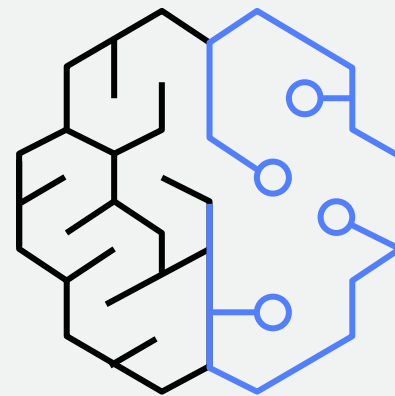
機械学習モデルを構築・運用・スケールする

Building

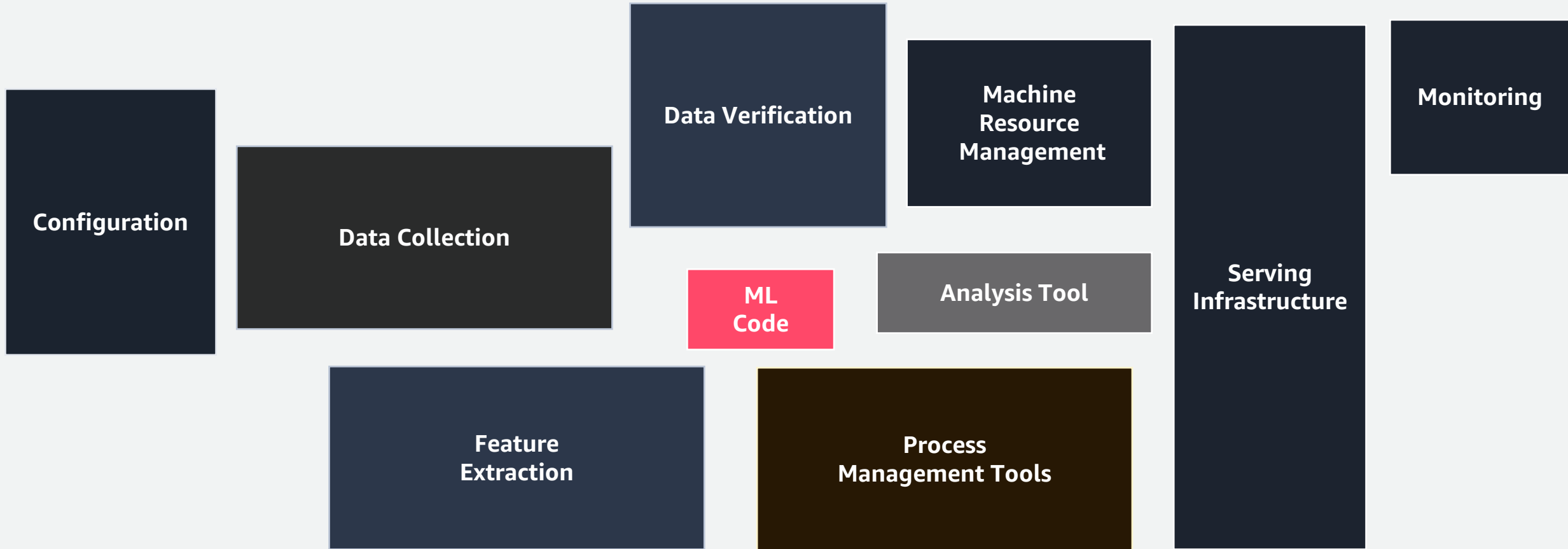
Deploying

Scaling

SageMaker



戦略2: クラウドの力を活用

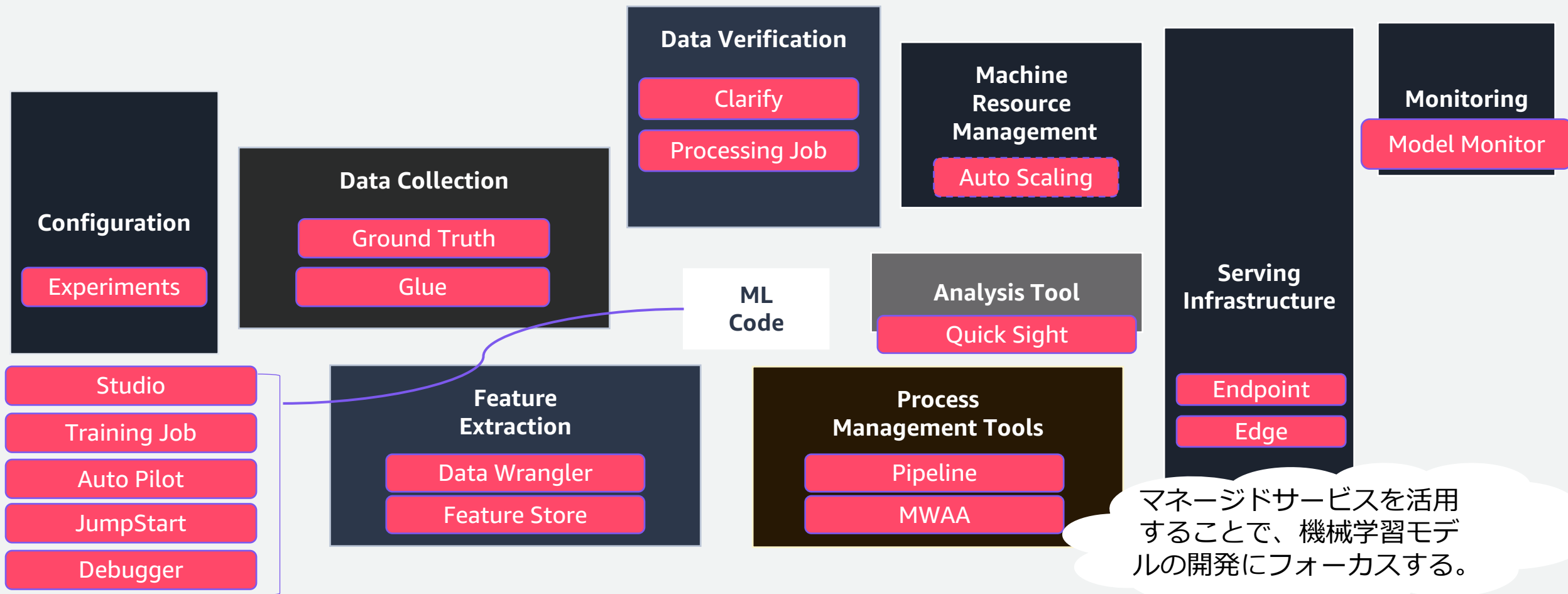


“Only a small fraction of real-world ML systems is composed of the ML code”

source: Hidden Technical Debt in Machine Learning Systems [D. Sculley, & al.] – 2015

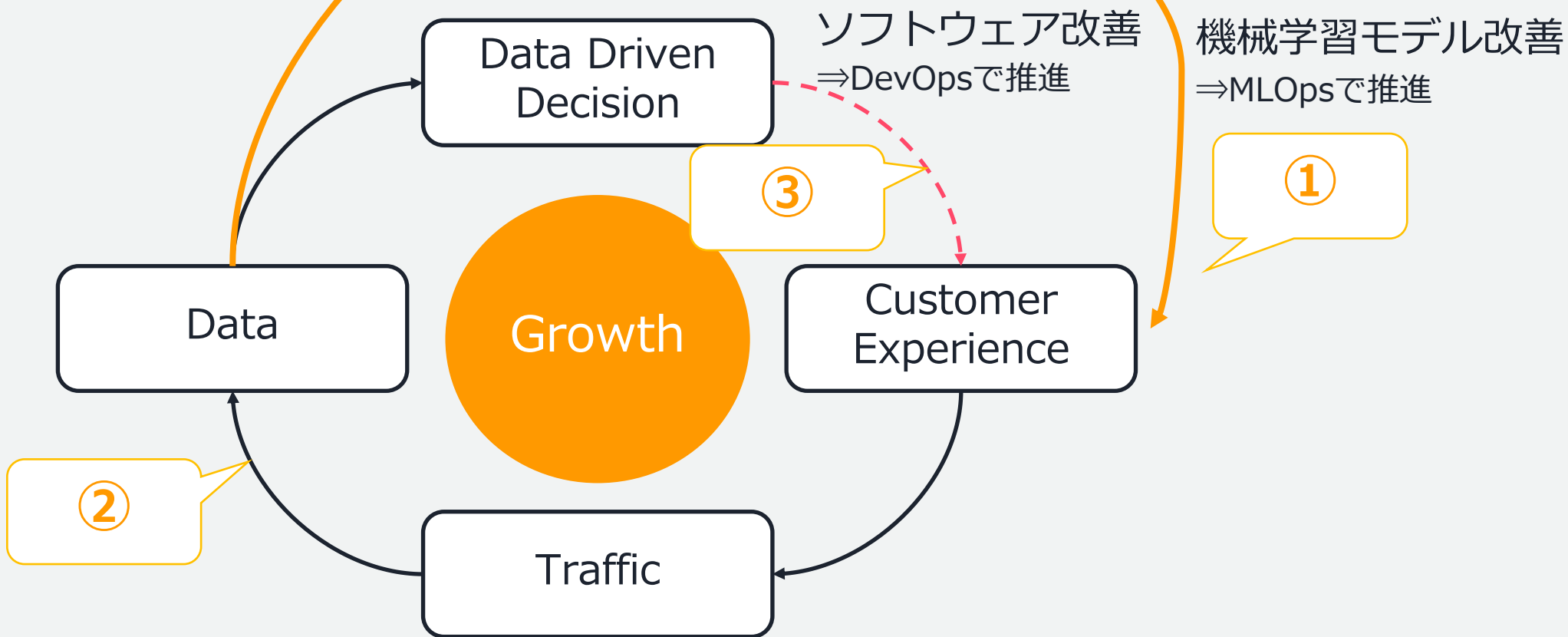
<https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf>

SageMakerはML Codeに集中したい開発者のために、機械学習の技術的負債を解消するマネージドサービスを提供します

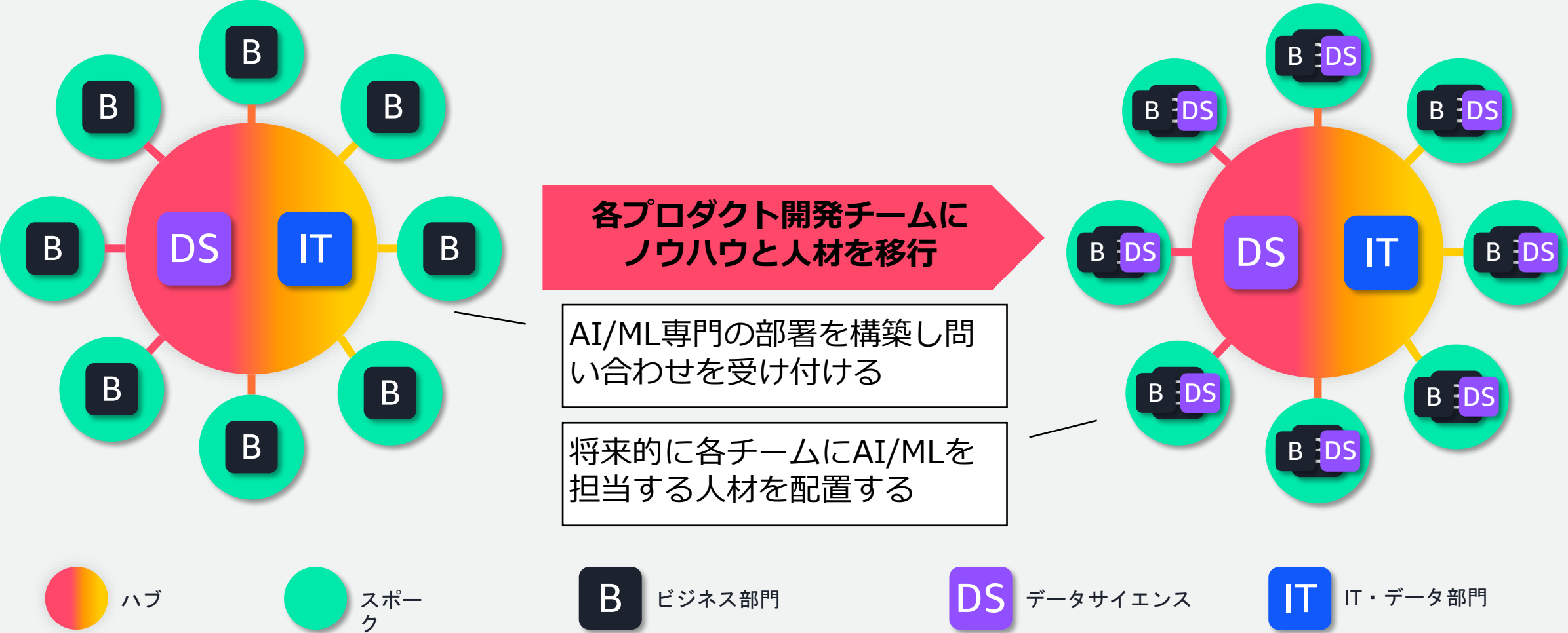


戦略2: POCを越えるスケーラビリティ

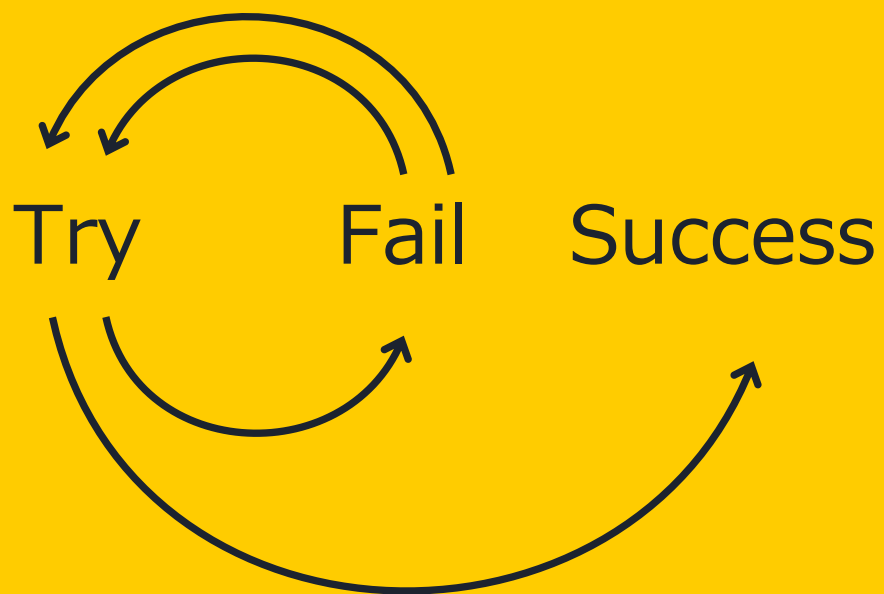
- ①機械学習がCustomer Experienceを改善しユーザー数が増える。
- ②増えたデータが学習データの源泉となる。
- ③モデルの改善速度と精度が、ソフトウェアによる改善を上回る。



戦略3: カルチャーシフトのための組織作り



戦略3: 前向きな失敗を奨励するカルチャー

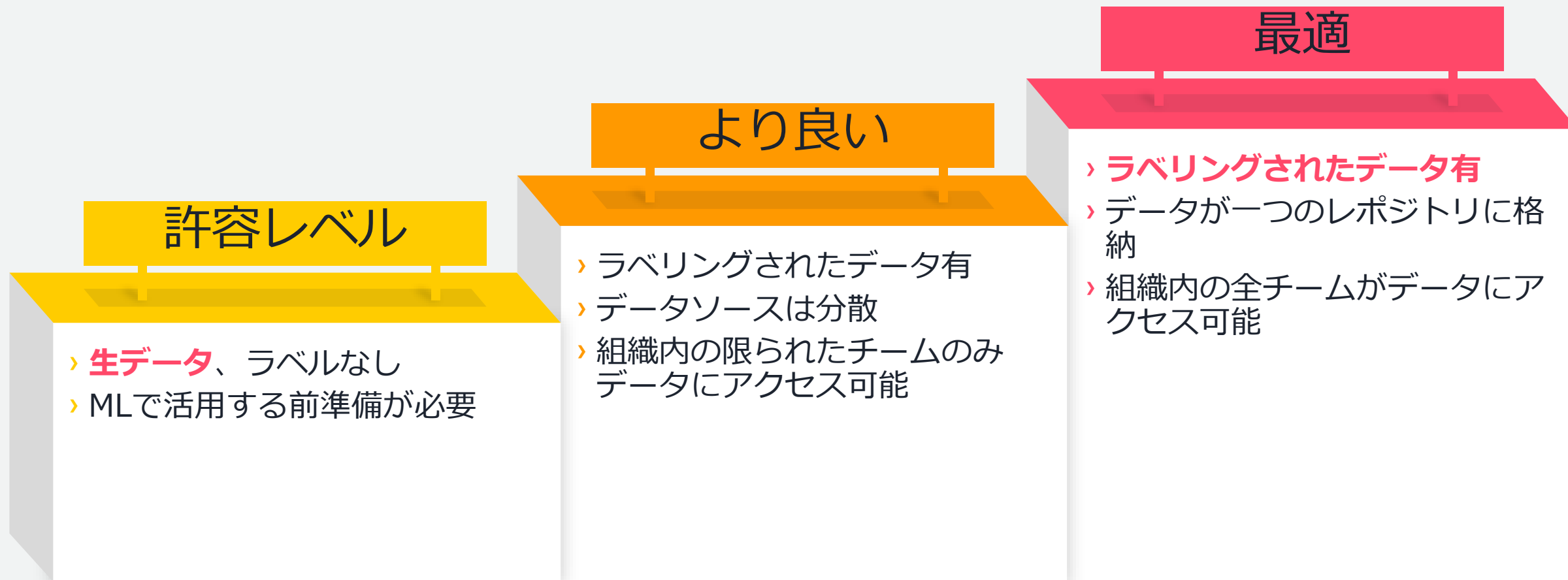


MLは反復が必要なプロセス

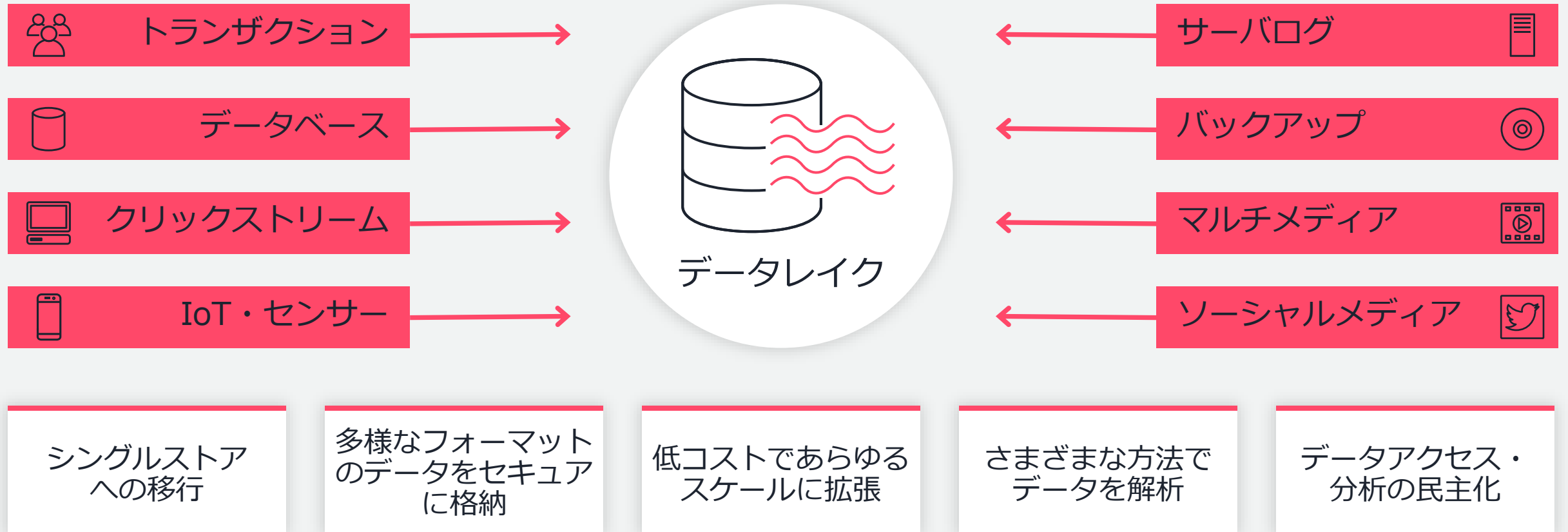
試行錯誤の繰り返しから学び、
失敗への耐性を身につける

1. 堅牢なAI戦略
2. データ戦略
3. 組織内のコラボレーション
4. 推進のためのリーダーシップ

データ戦略の推進：データの現状評価



データレイクを真実の情報源として活用



機械学習用データの要件を確認するための質問

活用できる?



必要なデータが
存在する



簡単にアクセス
可能

活用に課題なし?

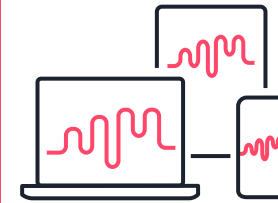


プライバシー
への配慮



セキュア

データの質は?



課題との関連性、
更新頻度



代表的かつ
偏りが無い

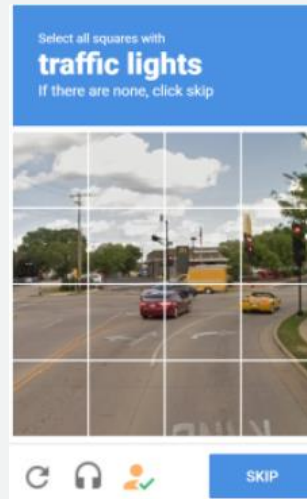
PoCの段階からデータの収集・活用プランを設計することで最適なフェーズへいち早く移行する。



PoCの段階からデータの収集・活用プランを作る

Module2の復習

- ①機械学習によるCustomer Experienceの改善が進むほど、収益増につながる仕組みを作れている(スケーラビリティ/パーソナライゼーション)。
- ②サービスの機能や運用を通じ、学習データが半自動的に蓄積される仕組みが構築されている(即時的なレスポンス)。
- ③短期的に費用対効果の高いルールベースによる精度改善が、限界に達しつつあること(複雑なロジック)。



👍 いいね!

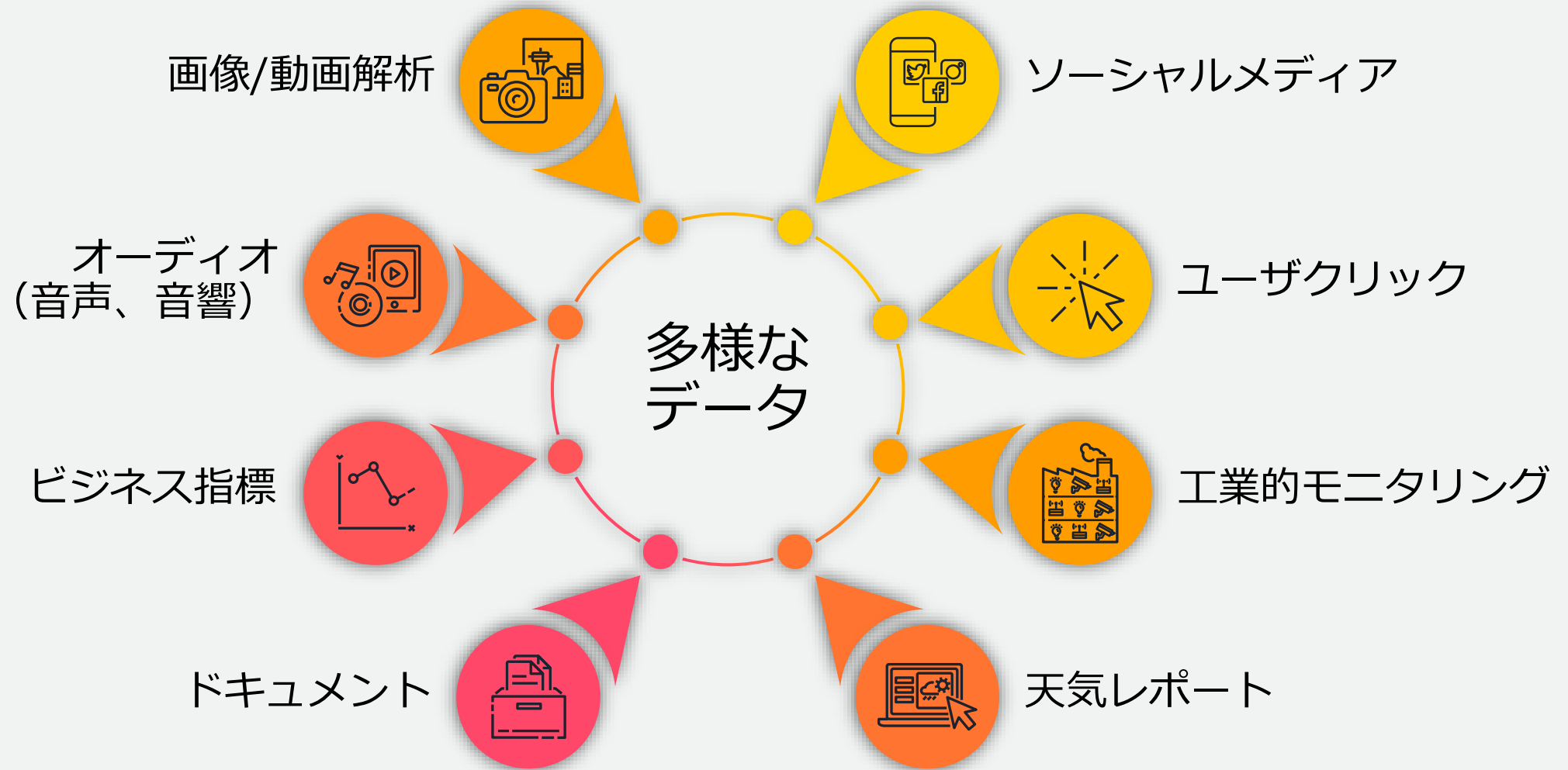
【パネル選択による認証】

ラベルデータを生成する仕組みをアプリケーションの機能にすることでデータ作成のコストを削減できる。

【いいねボタン】

機械学習による提案に対し人間がフィードバックを行う機能を持つことで再学習のタイミングを検知できる。

要件を満たすデータの収集



AWS Data Exchangeでは様々なデータを購入可能です。
販売頂くこともできます。

AWS Data Exchange

クラウド内でのサードパーティデータの容易な検索、購読、および使用

3,000 以上のサードパーティのデータセットを
閲覧する



豊富なデータセットカタログ



AWS との統合によるデータテクノロジーの向上



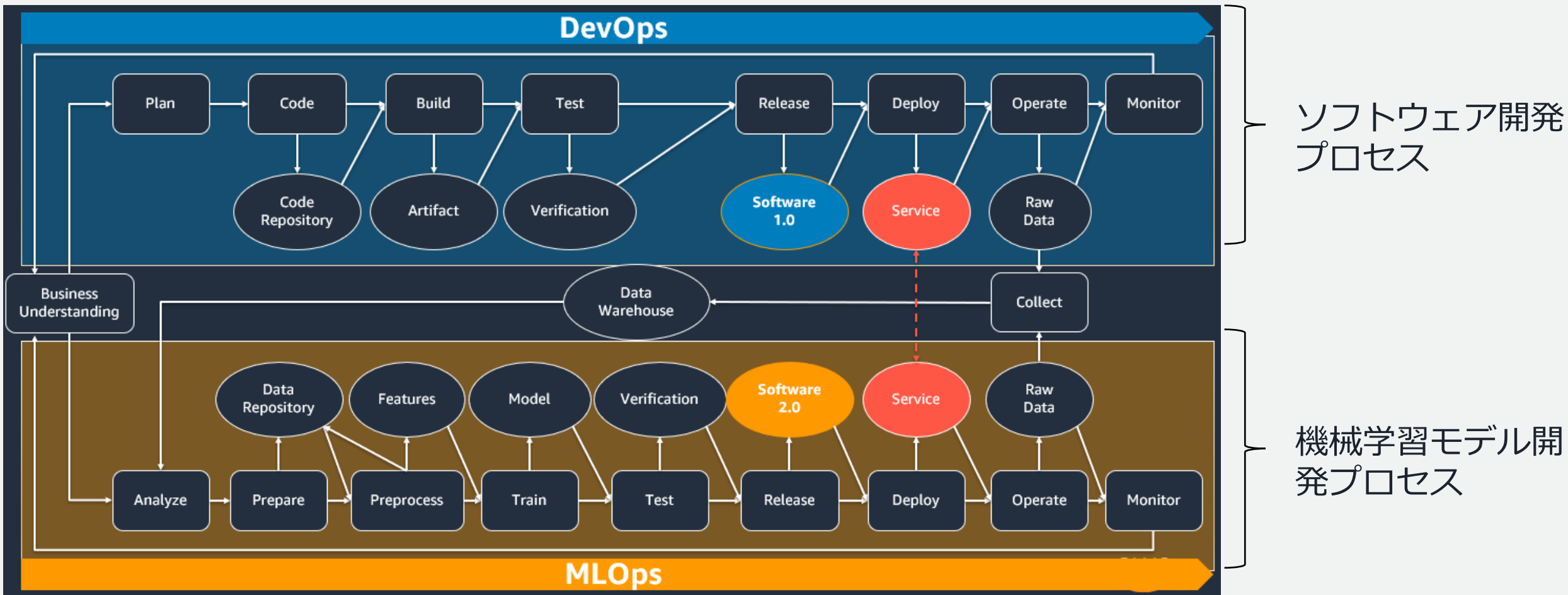
データ調達とガバナンスの合理化



データファイル、テーブル、API
を簡単に利用する

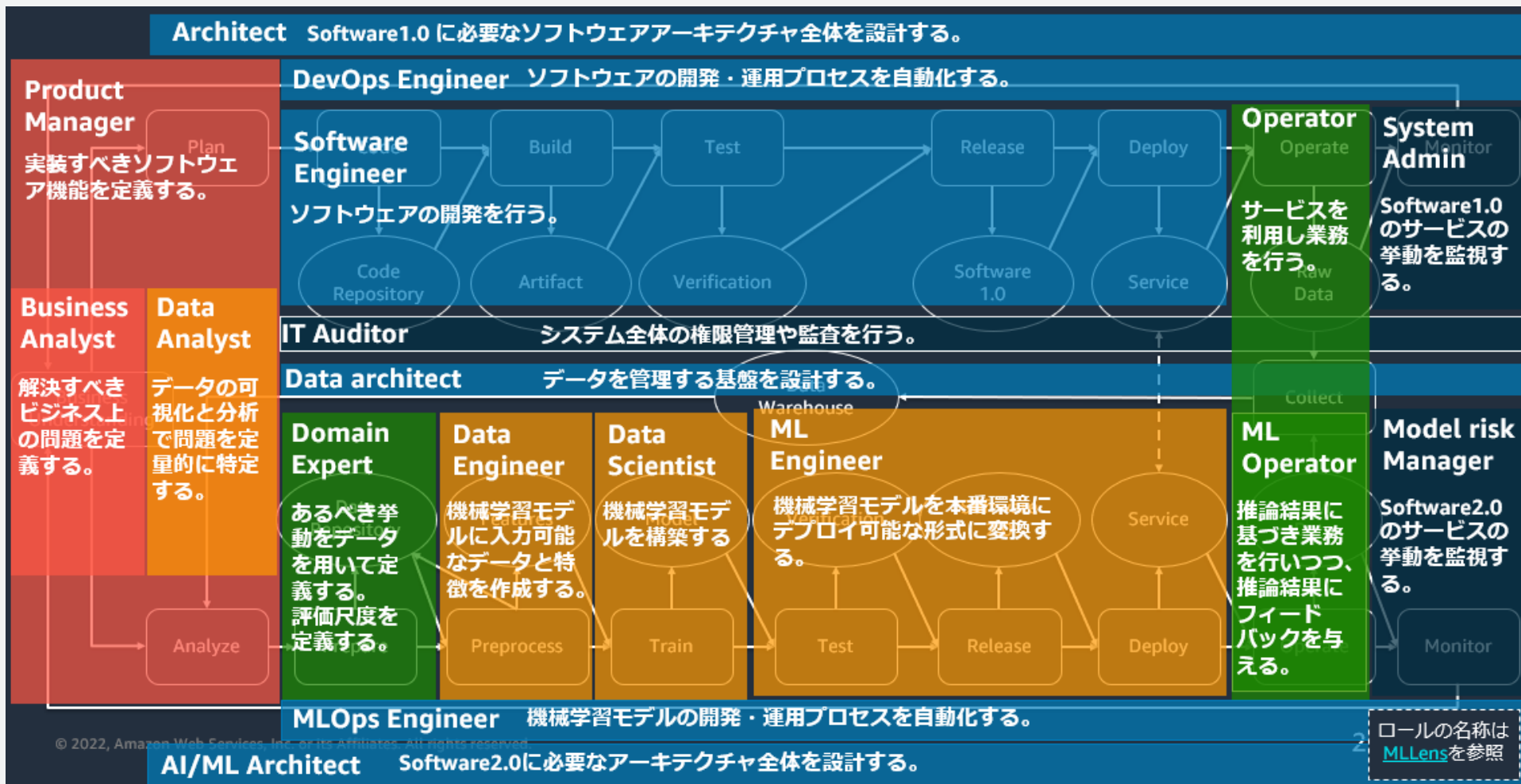
1. 堅牢なAI戦略
2. データ戦略
3. 組織内のコラボレーション
4. 推進のためのリーダーシップ

プロダクトに機械学習を組み込む場合の開発プロセス



<https://speakerdeck.com/icoxfog417/mlopsfalsekoremadetokorekara>

プロダクトに機械学習を組み込むには18の役割がかかわる



ソフトウェア開発
チーム

機械学習モデル開発
チーム

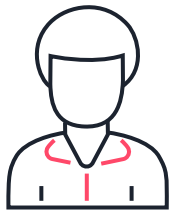
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

<https://speakerdeck.com/icoxfog417/mlopsfalsekoremadetokorekara>



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

パイロット機械学習モデル開発チームの一例



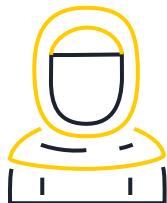
プロダクトマネージャー

役割:

MLプロジェクトがビジネス・テクノロジー双方の要件を満たし、期限内にデリバリーされることを確認

スキル:

プロジェクトマネジメント、リーダーシップ、MLアルゴリズムの概要の理解



データサイエンティスト

役割:

MLモデルの構築

スキル:

数学、統計、MLアルゴリズム、データ処理



MLエンジニア

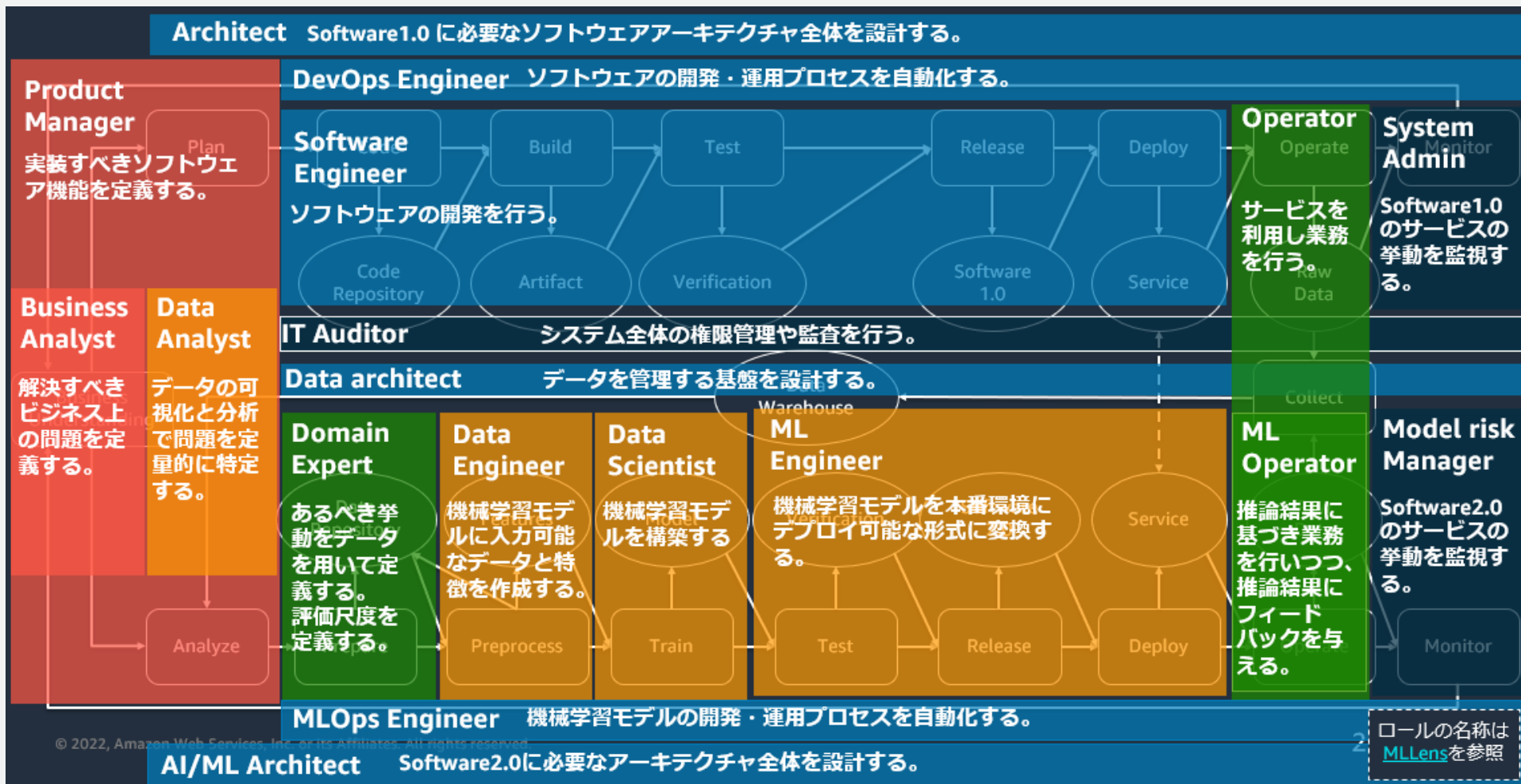
役割:

MLモデルのプロダクション化

スキル:

(MLアルゴリズム)、データパイプラインツール、アーキテクチャ設計、ソフトウェア開発

データサイエンティスト・MLエンジニアの参画



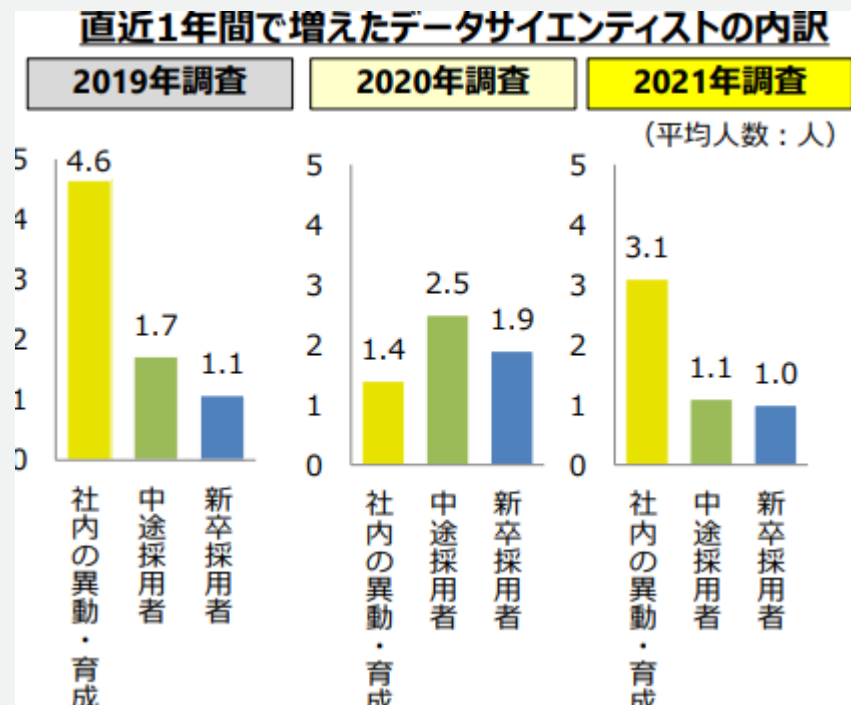
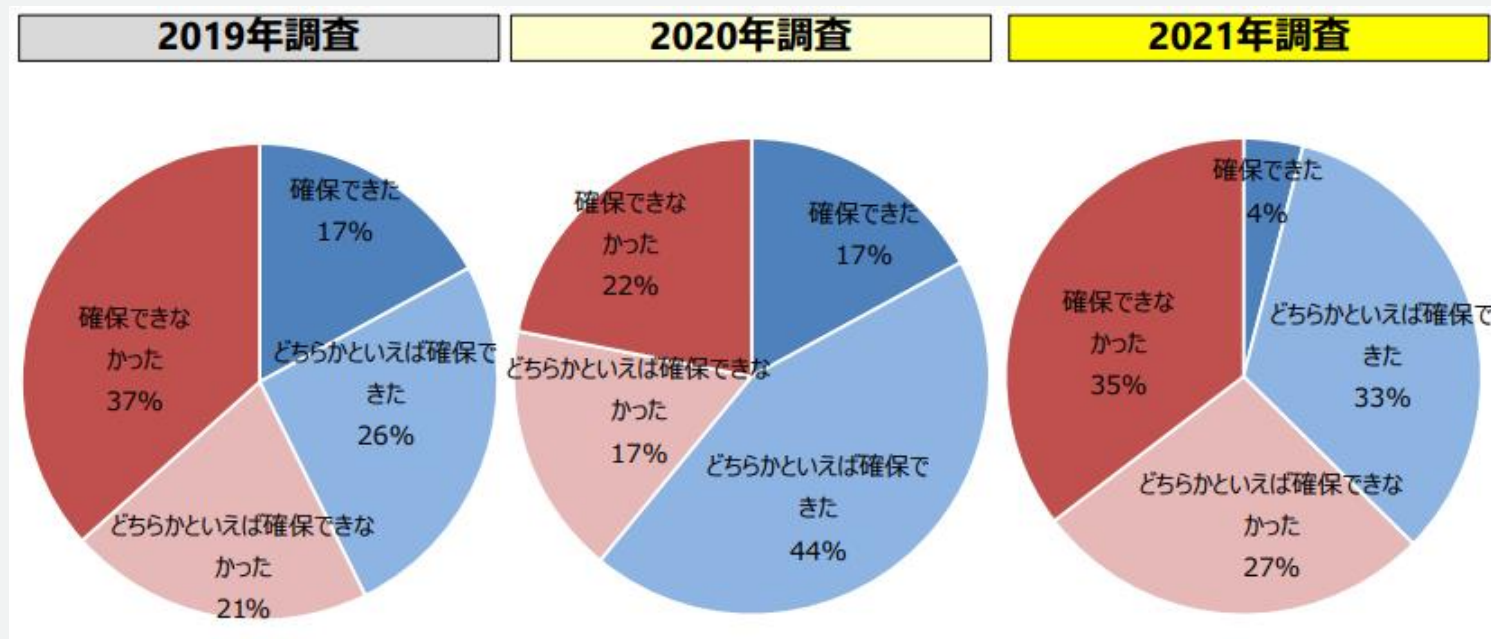
ソフトウェア開発
チーム

スキル獲得し移動

機械学習モデル開発
チーム

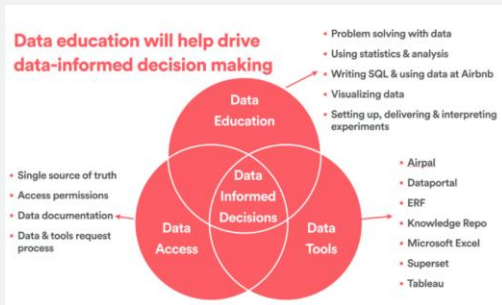
データサイエンティストのスキルセットの典型例

データサイエンティスト協会の2021年の調査結果では、62%が目標採用人数を確保できないと回答。社内での異動・育成で補っている。



出典:データサイエンティストの採用に関するアンケート (2022年3月31日)

機械学習のトレーニング: 先進的企業の事例



画像引用: リンク先記事より

[How Airbnb Democratizes Data Science With Data University](#)

データを「顧客の声」と定義し、全社員が顧客の声に基づいた意思決定ができるよう社内でのデータサイエンス教育を開始。半年で500名が受講。社内のデータを使用してトレーニングするため、30名ほどの社員がボランティアで講師を行っている。

[Scaling Machine Learning at Uber with Michelangelo](#)

技術系の採用者全員に、1週目に機械学習とUberの機械学習基盤であるMichelangeloの講義を実施。その後も、Michelangeloのアップデートに関する講習やオフィスアワーなどを実施し利用を支援。



画像引用: リンク先記事より

Airbnbは2017年、Uberは2018年の記事。Airbnbは2017時点でFacebookとDropboxにデータ活用で後れを取っていると述べているので、データ活用の先進企業はかなり先をいっている。

AWSが提供する機械学習のトレーニング

社内でのテクニカルスキル育成に投資しよう

ハンズオン学習

トレーニング



1. 堅牢なAI戦略
2. データ戦略
3. 組織内のコラボレーション
4. 推進のためのリーダーシップ

カギとなる質問



全てのチーム向け

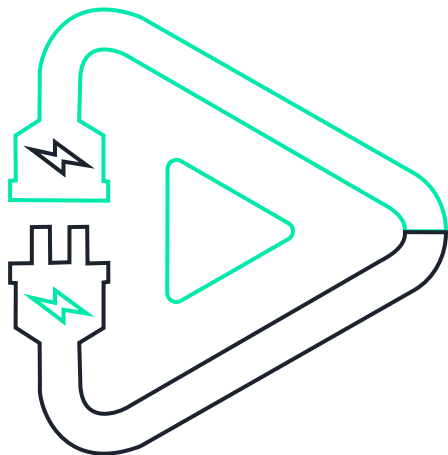
研究、ソフト開発、人事チームなども含めて
この質問を尋ねる

「どのように機械学習を使いますか」

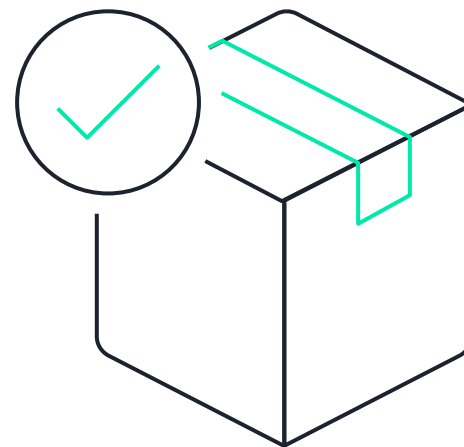
- › この質問への回答のために、各チームがビジネス・技術のエキスパートと協業することとなる

「わかりません」「使いません」
という回答は受け入れない

よくある間違い



AIを即効性のあるプラグ&プレイ技術としてとらえる



AIの活用範囲を狭く考えすぎる

リーダーシップのシミュレーション



考えてみましょう

あなたは、これからMLジャーニーを始めよう
としている小売業の会社のCXOです

既にアナリティクスのパイプラインを構築し
ていますが、MLを活用したさまざまなユース
ケースを実現したいと考えています

どのように始めますか？

MLジャーニーの開始

ML Journey

Learning

Experimenting

Prototyping

Building

Deploying

Scaling



興味はあるが、
実経験が浅い

堅牢なAI戦略

- › **重要度の高い**プロジェクトからPoCを始める
- › これらのプロジェクトを元にモメンタムを作る
- › AIへのシフトを始める理由を説明する

データ戦略

- › 既にあるデータを探索し、ドキュメント化する
- › そのデータがAIプロジェクトに使えるかを文書で記す
- › **データパイプラインがない場合**、構築のためのタイムラインを準備する

組織内コラボレーション

- › **各チームにデータの探索と実験を許可する**
- › 社外リソースと協業する
- › マネージドサービスを活用する

MLジャーニーの開始

ML Journey

Learning

Experimenting

Prototyping

Building

Deploying

Scaling



少数の実用化
実績

堅牢なAI戦略

- › 組織全体でAIケーパビリティを構築する
- › **PoCからプロダクションに移行する戦略を立てる**
- › チームを横断したコラボレーションを奨励する

データ戦略

- › **複数のデータソースを一つの「真実の情報源」（データレイク、データウェアハウスなど）に統合する戦略を立てる**
- › 組織全体のデータリテラシーを高める

組織内コラボレーション

- › **社内AIチームを結成する**
- › 既存のワークフォースを採用し、オンラインや対面型のコースでトレーニングする

MLジャーニーの開始

ML Journey

Learning

Experimenting

Prototyping

Building

Deploying

Scaling



複数の実用化実績により
ビジネスインパクトを創出

堅牢なAI戦略

- › **ビジネスの中核に長期的なAI戦略を組み込む**
- › 長期ビジョンを維持するためのAIリーダーシップ（支援体制）を作る
- › 人材採用のパイプラインを作る

データ戦略

- › 活用可能な**外部データを取り込み**、自社データの質と有効性を高める

組織内コラボレーション

- › **さまざまな社内チームと協業できるAI組織を作る**
- › データに対する探索と実験を継続する



ナレッジチェック: AI導入の支援

以下のうち、ML導入促進の初期段階において、リーダーがとるべきではないアクションはどれ？

- a) 実験と失敗の許容
- b) データの完全性およびアクセス性の確保
- c) 完全なMLインフラストラクチャの構築
- d) 最適なユースケースを特定するための素早いPoC作り
- e) 学際的なコラボレーションの促進

ナレッジチェック: AI導入の支援

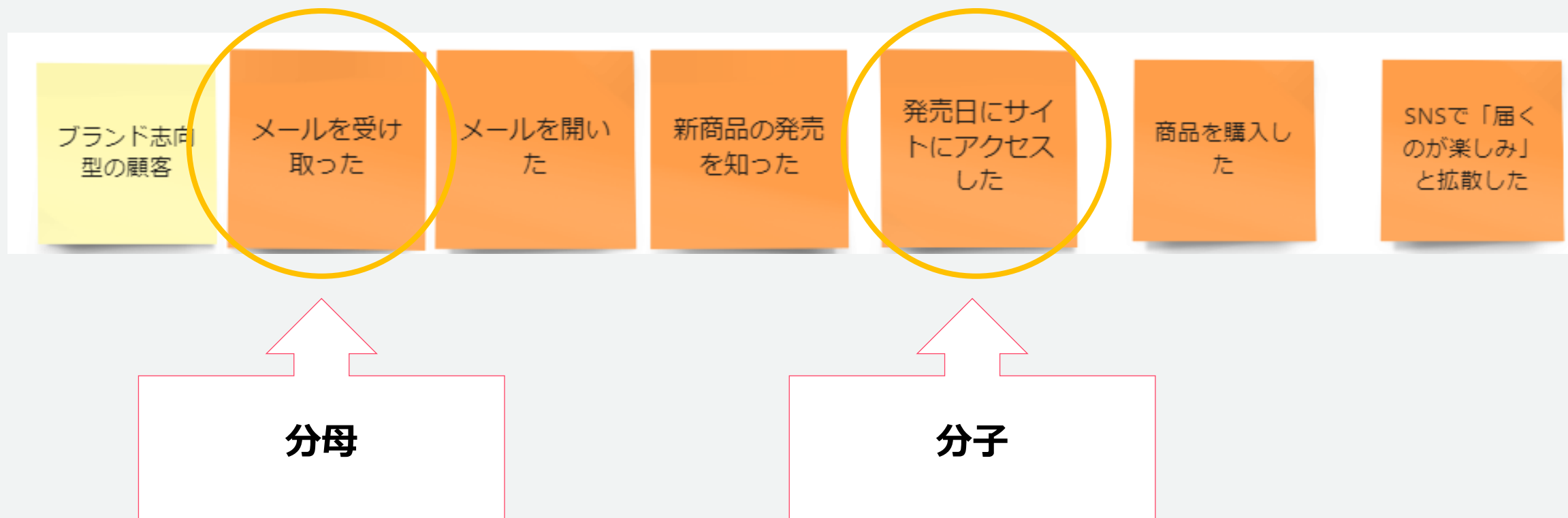
以下のうち、ML導入促進の初期段階において、リーダーがとるべきではないアクションはどれ？

- a) 実験と失敗の許容
- b) データの完全性およびアクセス性の確保
- c) 完全なMLインフラストラクチャの構築
- d) 最適なユースケースを特定するための素早いPoC作り
- e) 学際的なコラボレーションの促進

实践

実践

- 「応用」フェーズで作成したEvent Stormingから、顧客の体験を計測する方法を特定してください。



実践

- チームで機械学習に至るまでのステップを検討してください。
- 各ステップで必要なものを洗い出してください
 - 意思決定者の承認
 - 工数
 - コスト
 - データ

実践

- 最初のステップの実施を計画してください
 - 実施時期
 - Next Stepに進む/やめるためのクライテリア
 - AWSに依頼する支援
- CXOへの計画説明日をスケジュールしてください。

ML Enablement Workshop

振り返り



理解編：機械学習の入門

Q: なぜ機械学習に取り組む必要があるのか？

A: 顧客体験を迅速に改善するため。

機械学習の勝ちパターン

顧客体験の改善が利用者の増加、利用者の増加がデータの増加、データの増加がモデルの精度向上、モデルの精度向上が顧客体験の改善につながるサイクルが回る。

機械学習の特性

データから帰納的に処理を実装する技術。複雑なロジックを高精度で解けるなどのメリットがあるが、説明性や精度100%が求められる状況には向かない。

応用編:機械学習の勝ちパターンを応用する

Q: 事例通りにしても効果が出ないのはなぜか?

A: 顧客の問題がプロダクトごと異なるから。

良いビジネス課題

良いビジネス課題はビジネスの問題と顧客の問題を同時に解決する。Event Stormingをはじめとしたストーリーを理解する手法は、顧客の問題が解けているか確認するのに使える。

優先順位付け

緊急度と重要度の観点から、課題の優先順位をつけることができる。

開始編: データに基づく改善を開始する

Q: 機械学習の実装がはじまらないのはなぜ?

A: 不確実な成果と確実なコストの比較結果

機械学習の価値は、モデルを開発しなくても計測できる

データに基づく手動の業務、演繹的なロジックの実装など段階的なステップを踏むことで効果を実感しながら機械学習の導入に至れる。

機械学習をスケールさせるための組織作り

全社的なAI戦略・データ戦略・組織戦略の要件とそれらを推進するためのリーダーシップについて確認。



Thank you!

This material is based on the AWS Machine Learning Embark Program,
an immersive ML training program provided by Amazon Machine Learning Solutions Lab