

Professional Machine Learning Engineer

認定試験ガイド

Professional Machine Learning Engineer は、Google Cloud の機能と従来の ML アプローチに関する知識を生かして、AI ソリューションを構築、評価、製品化、最適化します。ML エンジニアは、大規模で複雑なデータセットを処理して、再現可能かつ再利用可能なコードを作成します。ML エンジニアは、基盤モデルに基づいて生成 AI ソリューションを設計、運用します。ML エンジニアは、責任ある AI について検討し、他の職務と密接に連携して、AI ベースのアプリケーションの長期的な成功を確実にします。ML エンジニアには、優れたプログラミング スキルと、データプラットフォームや分散データ処理ツールに関する経験があります。ML エンジニアは、モデル アーキテクチャ、データと ML パイプラインの作成、生成 AI、指標の解釈に精通しています。ML エンジニアは、MLOps、アプリケーション開発、インフラストラクチャ管理、データ エンジニアリング、データ ガバナンスの基本コンセプトに精通しています。ML エンジニアは、組織全体のチームが AI ソリューションを使用できるようにします。ML エンジニアは、モデルのトレーニング、再トレーニング、デプロイ、スケジューリング、モニタリング、改善を行い、スケーラブルでパフォーマンスの高いソリューションを設計、作成します。

セクション 1: ローコード AI ソリューションの構築 (試験内容の 13%)

1.1 BigQuery ML を使用した ML モデルの開発。考慮事項:

- ビジネス上の問題に基づく適切な BigQuery ML モデル(線形分類、バイナリ分類、回帰、時系列、行列分解、ブーストツリー、オートエンコーダなど)の構築
- BigQuery ML を使用した特徴量エンジニアリングや特徴選択
- BigQuery ML による予測の生成

1.2 ML API または基盤モデルを使用した AI ソリューションの構築。考慮事項:

- Model Garden の ML API を使用したアプリケーションの構築
- 業界固有の API を使用したアプリケーションの構築(Document AI API、Retail API など)
- Vertex AI Agent Builder を使用した検索拡張生成(RAG)アプリケーションの実装

1.3 AutoML を使用したモデルのトレーニング。考慮事項:

- AutoML 用のデータの準備(特徴選択、データラベル付け、AutoML での表形式ワークフローなど)
- 利用可能なデータ(表形式、テキスト、音声、画像、動画など)を使用したカスタムモデルを使用

Google Cloud

- 表形式データに AutoML を使用
- AutoML を使用した予測モデルの作成
- トレーニング済みモデルの構成とデバッグ

セクション 2: チーム内およびチーム間の連携によるデータとモデルの管理(試験内容の 14% 以下)

2.1 組織全体のデータの探索と前処理 (Cloud Storage、BigQuery、Spanner、Cloud SQL、Apache Spark、Apache Hadoop など)。考慮事項:

- 異なるタイプのデータ(表形式、テキスト、音声、画像、動画など)の効率的なトレーニングを目的とする整理
- Vertex AI でのデータセットの管理
- データの前処理 (Dataflow、TensorFlow Extended [TFX]、BigQuery など)
- Vertex AI Feature Store での特徴の作成と統合
- データの使用や収集に関するプライバシーの影響(個人情報 (PII) や保護対象保健情報 (PHI))といった機密データの処理など)
- 推論のための Vertex AI へのさまざまなデータソース(テキストドキュメントなど)の取り込み

2.2 Jupyter ノートブックを使用したモデルのプロトタイプング。考慮事項:

- Google Cloud 上での適切な Jupyter バックエンドの選択 (Vertex AI Workbench、Colab Enterprise、Dataproc 上のノートブックなど)
- Vertex AI Workbench におけるセキュリティに関するベストプラクティスの適用
- Spark カーネルの使用
- コードソース リポジトリの統合
- Vertex AI Workbench で一般的なフレームワーク (TensorFlow、PyTorch、sklearn、Spark、JAX など)を使用した Vertex AI Workbench でのモデル開発
- Model Garden のさまざまな基盤モデルとオープンソース モデルの活用

2.3 ML テストのトラッキングと実行。考慮事項:

- 開発とテストに適した Google Cloud 環境 (Vertex AI Experiments、Kubeflow Pipelines、TensorFlow と PyTorch を使用した Vertex AI TensorBoard など)をフレームワークに応じて選択
- 生成 AI ソリューションの評価

セクション 3: プロトタイプ of ML モデルへのスケーリング (試験内容の 18%)

3.1 モデルの構築。考慮事項:

- ML フレームワークとモデル アーキテクチャの選択
- 解釈可能性要件のあるモデル手法

3.2 モデルのトレーニング。考慮事項:

- Google Cloud (Cloud Storage、BigQuery など) 上のトレーニング データ (表形式、テキスト、音声、画像、動画など) を整理する
- さまざまな種類のファイル (CSV、JSON、画像、Hadoop、データベースなど) をトレーニング に取り込む
- さまざまな SDK を使用したモデルのトレーニング (Vertex AI カスタムトレーニング、Google Kubernetes Engine 上の Kubeflow、AutoML、表形式のワークフローなど)
- 分散トレーニングによる信頼性の高いパイプラインの組織化
- ハイパーパラメータ調整
- ML モデルのトレーニング失敗のトラブルシューティング
- 基盤モデル (Vertex AI Model Garden など) のファイン チューニング

3.3 トレーニングに適したハードウェアの選択。考慮事項:

- コンピューティング オプションとアクセラレータ オプションの評価 (CPU、GPU、TPU、エッジデバイスなど)
- TPU と GPU を使用した分散トレーニング (Vertex AI 上の Reduction Server、Horovod など)

セクション 4: モデルのサービングとスケーリング (試験内容の 20%)

4.1 モデルのサービング。考慮事項:

- バッチ推論とオンライン推論 (Vertex AI、Dataflow、BigQuery ML、Dataproc など)
- さまざまなフレームワーク (PyTorch、XGBoost など) を使用したモデルのサービング
- Model Registry でのモデルの整理
- 1 つのモデルの異なるバージョンを使用した A/B テスト

4.2 オンライン モデル サービングのスケーリング。考慮事項:

- Vertex AI Feature Store を使用した特徴の管理とサービング
- パブリック エンドポイントとプライベート エンドポイントへのモデルのデプロイ

Google Cloud

- 適切なハードウェアの選択 (CPU、GPU、TPU、エッジなど)
- スループットに基づいたサービング バックエンドのスケーリング (Vertex AI Prediction、コンテナ化されたサービングなど)
- 本番環境でのトレーニングとサービングのための ML モデルの調整 (簡素化手法、パフォーマンス、レイテンシ、メモリ、スループット向上のための ML ソリューションの最適化など)

セクション 5: ML パイプラインの自動化とオーケストレーション (試験内容の 22%)

5.1 エンドツーエンドの ML パイプラインの開発。考慮事項:

- データとモデルの検証
- トレーニングとサービングの間で一貫したデータ前処理の保証
- Google Cloud でのサードパーティ パイプラインのホスティング (MLflow など)
- コンポーネント、パラメータ、トリガー、コンピューティングのニーズの特定 (Cloud Build、Cloud Run など)
- オーケストレーション フレームワーク (Kubeflow Pipelines、Vertex AI Pipelines、Cloud Composer など)
- ハイブリッド戦略またはマルチクラウド戦略
- TFX コンポーネントまたは Kubeflow DSL を使用したシステムの設計 (Dataflow など)

5.2 モデルの再トレーニングの自動化。考慮事項:

- 適切な再トレーニング ポリシーの決定
- 継続的インテグレーションと継続的デリバリー (CI / CD) パイプライン (Cloud Build、Jenkins など) を使用したモデルのデプロイ

5.3 メタデータのトラッキングと監査。考慮事項:

- モデルのアーティファクトとバージョンの追跡と比較 (Vertex AI Experiments、Vertex ML Metadata など)
- モデルおよびデータセットのバージョンの指定
- モデルとデータ系列

セクション 6: AI ソリューションのモニタリング (試験内容の 13%)

6.1 AI ソリューションに対するリスクの特定。考慮事項:

- データやモデルの意図しない搾取 (ハッキングなど) からの保護による、安全な AI システムの構築
- Google の責任ある AI への取り組みとのすり合わせ (バイアスのモニタリングなど)
- AI ソリューションの準備状況の評価 (公平性、バイアスなど)
- Vertex AI でのモデルの説明可能性 (Vertex AI Prediction など)

6.2 AI ソリューションのモニタリング、テスト、トラブルシューティング。考慮事項:

- 継続的な評価指標の確立 (Vertex AI Model Monitoring、Explainable AI など)
- トレーニング / サービング スキューのモニタリング
- 特徴アトリビューションのドリフトのモニタリング
- ベースライン、シンプルなモデル、時間枠に対するモデルのパフォーマンスのモニタリング
- 一般的なトレーニング エラーやサービング エラーのモニタリング