

プロフェッショナルデジタルスキル（データ・AI）試験（仮称）

科目 B のサンプル問題

問1 A社は、全社データ基盤において、中央のデータ管理部門にデータ管理の責任を集中させず、各部門がデータのオーナーシップをもつ分散型のデータマネジメント方針を採用している。

営業部門は、プロバイダー（提供者）として“顧客別月間売上データプロダクト”を公開している。A社では、データプロダクトにアクセスするためには、コンシューマー（利用者）がデータコントラクトの品質や仕様を確認し、システム上で利用合意の手続を行うルールとしている。コンシューマーであるマーケティング部門は、このデータプロダクトの仕様を定めたデータコントラクト（以下、データコントラクトv1という）に利用合意し、解約予測AIのインプットとして運用している。なお、A社は、データの乱立を防ぐために、データコントラクトの複数バージョンの並行公開を禁止している。マーケティング部門が利用しているデータコントラクトv1を図1に示す。

```
schema:
  - name: customer_id
    type: string
  - name: amount
    type: decimal
    description: "顧客の月間売上金額（単位：千円）"
    description_approved: true
quality_slo:
  completeness:
    - name: customer_id
      allow_null: false
```

注記 YAML形式で記述している。

図1 データコントラクトv1（抜粋）

ある日、営業部門は、データコントラクトv1を参照した経営企画部門から、“新規に利用したいが、ダッシュボード用に円単位のデータがほしい”という要望を受けた。

この要望を受けて、営業部門は、新規で円単位のカラムを用意することも検討したが、既存の物理名（amount）の一般的な意味合いに合わせる方がよいと独自の判断をして、データ基盤上の実テーブルの既存カラムの値を円単位に変更する方針にした。

また、既存の仕様を担保するために、千円単位の元のデータを格納する amount_k カラムを実テーブルに新たに追加し、データコントラクトも新バージョン（以下、データコントラクト v2 という）に更新した。この際、営業部門は、amount を円単位に変更した旨を経営企画部門に伝えたが、マーケティング部門には伝えていなかった。

A 社では、データコントラクトを更新する際、AI がカラム名と実データの数値から推測して説明文（description）を自動生成する補助ツールを導入している。ここで、人間が一度確定させたデータ定義を AI が誤って上書きしてしまうことを防ぐために、未承認（description_approved: false）に設定されたカラムだけを生成対象とし、その結果をプロバイダーが確認・修正した後に承認（description_approved: true）とする仕組みとしている。今回のデータコントラクトの修正時には amount カラムの description_approved は true のまま変更せずに、amount_k カラムだけ説明文を AI で自動生成した。営業部門が生成された説明文を確認して承認した後の、データコントラクト v2 を図 2 に示す。

```
(省略)
- name: amount
  type: decimal
  description: "顧客の月間売上金額（千円単位）"
  description_approved: true
- name: amount_k
  type: decimal
  description: "顧客の月間売上金額（千円単位）"
  description_approved: true
(省略)
```

注記 YAML 形式で記述している。

図 2 データコントラクト v2（抜粋）

経営企画部門は、利用合意しようと図 2 を確認したとき、amount の説明文における単位が千円のままであることに気付いた。営業部門から amount の単位は円単位であると聞いていたが、自部門の利用としては問題ないと判断し、合意した。

後日、経営企画部門のダッシュボードは正常に稼働した。一方、マーケティング部門の解約予測 AI は顧客の売上げが突如 1000 倍になったと誤認し、優良顧客であると判定して、解約確率について極端に低い数値を予測するという異常を引き起こした。データマネジメント担当の B さんが調査した結果、この事象の原因として、a が特定された。

B さんは、本件における課題の解消のために、プロバイダーとコンシューマーが遵

守すべき新たなガイドラインを制定した。コンシューマー向けの３か条を次に示す。

第 1 条 新規契約時： b

第 2 条 契約変更時：利用合意中のデータコントラクトに変更の要望がある場合は、
プロバイダーに通知すること。ただし、プロバイダー主導で他
のコンシューマーと協議の上、変更可否が判断される。

第 3 条 契約破棄時：システムの廃止などによってデータの利用が不要になった場合
は、速やかに利用合意を解除し、プロバイダーに通知すること。

新たなガイドラインを周知して運用を徹底したことによって、データの仕様の不一致に起因する同様の事象は発生しなくなった。

設問 本文中の a , b に入れる字句の組合せはどれか。解答群の中から最も適切なものを選び。

- (一) 営業部門が既存データのセマンティクスを変更したにもかかわらず，説明文のメタデータの見直しを怠ったことと，後方互換性のない破壊的変更を既存利用者のマーケティング部門に通知して再合意を得なかったこと
- (二) 営業部門が既存データのセマンティクスを変更した際，AI の自動生成の補助ツールが未承認フラグのカラムだけを対象とし，テーブル全体のデータ定義を再スキャンする仕様になっていなかったため，営業部門が既存データに対する説明文のメタデータの見直し漏れを検知できなかったこと
- (三) 経営企画部門が既存のデータコントラクトに対して新たな要求を出したことが，amount の値を千円単位から円単位へ変更するという事象を引き起こし，マーケティング部門の解約予測 AI に影響を与えたこと
- (四) 実データをダウンロードして独自の統計分析ツールで単位や欠損の有無をチェックし，データコントラクトと差異がないことを確認して利用合意すること。
- (五) 提示されたデータコントラクトの定義に内容の矛盾や不備がある場合は，プロバイダーに修正を要求し，正しい定義が反映されてから利用合意すること。
- (六) データの仕様がデータコントラクトと異なっていた場合に発生する AI の誤予測などのビジネス損失について，事前にプロバイダーに報告すること。

解答群

	a	b
ア	(一)	(四)
イ	(一)	(五)
ウ	(一)	(六)
エ	(二)	(四)
オ	(二)	(五)
カ	(二)	(六)
キ	(三)	(四)
ク	(三)	(五)
ケ	(三)	(六)

問2 ある EC サイトの毎月の利益を最大化することを目標に、利益に関連するデータ間の関係性を確認したい。利益は次式で計算され、固定費は 1 か月当たり 10 (万円) とする。

$$\begin{aligned} \text{利益 (万円)} = & \text{売上 (万円)} - \text{訪問者数 (万人)} \times 5 \text{ (円)} \\ & - \text{SNS 広告費 (万円)} - \text{メール配信数 (万通)} \times 0.5 \text{ (円)} \\ & - \text{固定費 (万円)} \end{aligned}$$

〔データの可視化〕

利益に関連するデータ間の関係性を確認するために、散布図行列を用いて 24 か月分のデータを可視化した。利益、売上、訪問者数、SNS 広告費、メール配信数の各変数を用いて作成した散布図行列を図 1 に示す。散布図行列の非対角成分には各変数間の散布図が示されており、対角成分には各変数のヒストグラムが示されている。散布図行列から、a。

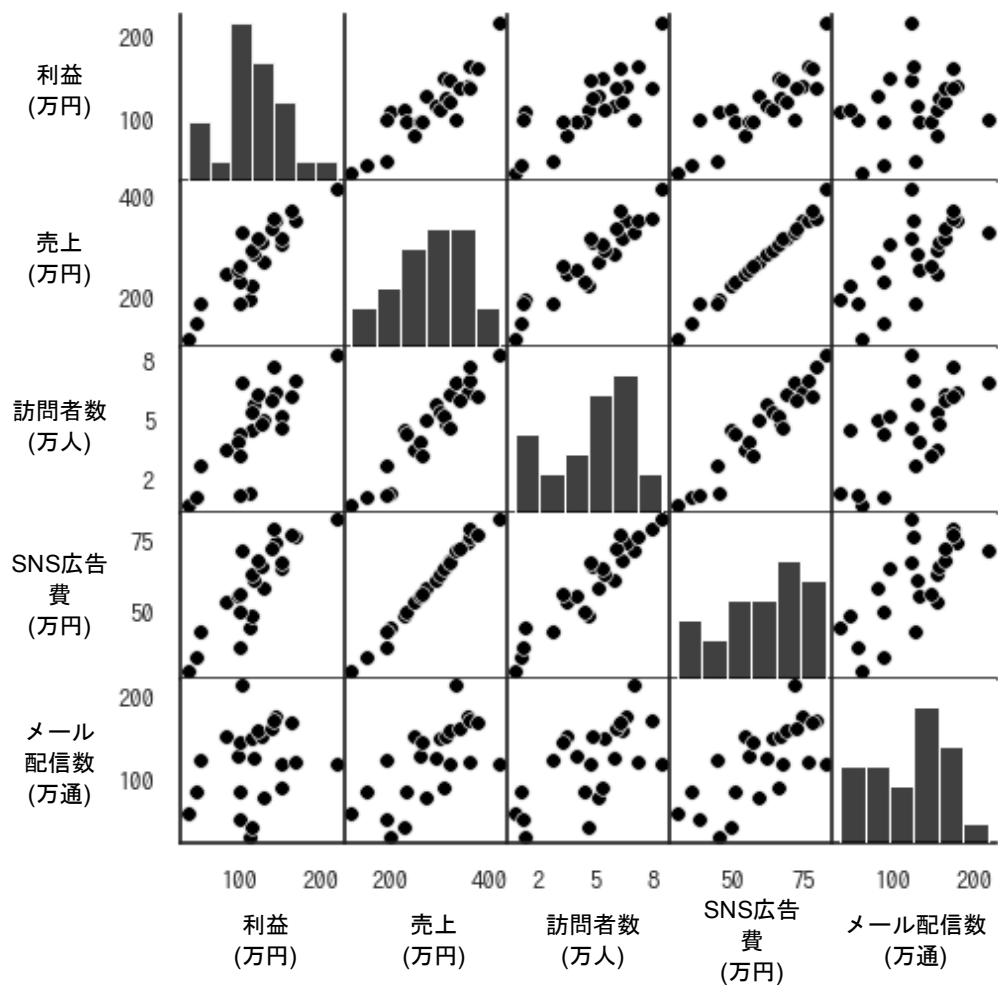


図 1 散布図行列

〔回帰分析〕

次に、EC サイト運営において調整対象の SNS 広告費とメール配信数を説明変数の候補，利益を目的変数の候補とし，幾つかのパターンで回帰分析をすることにした。モデル 1 では SNS 広告費を，モデル 2 ではメール配信数を，モデル 3 では SNS 広告費とメール配信数を説明変数にした回帰分析を行ったところ，表 1 に示す SNS 広告費の回帰係数，メール配信数の回帰係数，切片及び決定係数 R^2 を得た。回帰分析の結果から，決定係数 R^2 が最大であるモデル 3 が，利益の変動を最もよく説明でき，利益の予測に適したモデルであることが分かる。モデル 3 の計算結果によると，b といえる。

表 1 回帰分析の結果

モデル	SNS 広告費の係数	メール配信数の係数	切片	決定係数 R^2
モデル 1	2.334		−22.753	0.718
モデル 2		0.220	91.031	0.064
モデル 3	3.637	−0.586	−28.462	0.949

注記 網掛けはモデルの説明変数ではないことを示す。

設問 本文中の a , b に入れる字句の組合せはどれか。解答群の中から最も適切なものを選べ。

- (一) SNS 広告費と売上の間には、強い負の相関が見られる
- (二) メール配信数と利益の間には強い正の相関が見られ、利益がメール配信数に依存していることが分かる
- (三) SNS 広告費とメール配信数の間には弱い正の相関が見られ、共線性の存在が示唆されている
- (四) メール配信数が同じであれば、SNS 広告費を 1 万円増やすと利益は約 3.6 万円増えると予測できる
- (五) 切片がマイナスなので、このモデルでは利益が出ないことを示している
- (六) 決定係数 $R^2 = 0.949$ なので、このモデルの精度は 94.9%である

解答群

	a	b
ア	(一)	(四)
イ	(一)	(五)
ウ	(一)	(六)
エ	(二)	(四)
オ	(二)	(五)
カ	(二)	(六)
キ	(三)	(四)
ク	(三)	(五)
ケ	(三)	(六)

問3 Bさんは、精密機器メーカーであるA社の情報システム部門に所属している。A社は、保守の現場で発生するトラブルに関連する情報をエンジニアが探し出すのに時間がかかりすぎているという課題を抱えている。その原因としては、製品仕様書が頻繁に改訂されている上に、文章量が多いことが挙げられる。

そこで、Bさんは現場の保守点検を担うエンジニア向けに、大規模言語モデル（以下、LLM という）を活用した対話型のトラブルシューティング支援システム（以下、支援システムという）のプロトタイプを構築することになった。エンジニアや組織幹部へのヒアリングの結果、実装すべき要件とその理由を表1のとおりに定義した。

表1 要件とその理由

要件	理由
製品仕様書の改訂が発生した場合、速やかに支援システムに反映されること	製品仕様書の改訂は頻繁に起こるので、現場のエンジニアが古い情報を参照するおそれがあるから
支援システムの提示した解決策が、どの製品仕様書に基づいているかをエンジニアが直接確認できること	LLM による推論だけでもっともらしく回答されてしまうと、誤った作業を誘発するリスクがあるから

Bさんは、次の二つのアプローチを検討し、要件を満たせるかどうかを評価した。

〔検討したアプローチ〕

アプローチ①：現場の複雑な状況判断に対応するために、製品仕様書を用いて生成した学習データセットで追加学習したドメイン特化型のLLMを構築・運用する。

アプローチ②：追加学習は行わず、エンジニアの質問に関連する情報を検索してプロンプトに付加し、LLMは与えられた情報に基づいて回答を生成する。

Bさんは、二つのアプローチと要件を照らし合わせて検討した結果、製品仕様書の改訂をシステムに反映させる際に、アプローチ①に比べて、アプローチ②の方が短時間で済み、費用も低く抑えられることからアプローチ②を採用することにした。これを実現するために、aをプロンプトに付加し、LLMが与えられた情報に基づいて回答を生成する方式を設計した。この方式は、LLMのパラメータを更新せずに、提示された情報から回答を導き出すbという能力を活用したものである。また、支援システムが回答を出力するときにcことで、現場のエンジニアが回答の妥当性を評価し、安全に作業を進められる仕組みを構築できる。

設問 本文中の a ～ c に入れる字句の組合せはどれか。解答群の中から最も適切なものを選べ。

- (一) 外部データベースから検索・抽出された，質問の関連文書
- (二) LLM がもつ知識から情報を検索・抽出するための，ベクトル変換した質問文
- (三) パラメトリック知識
- (四) 継続的学習 (Continual Learning)
- (五) 文脈内学習 (In-Context Learning)
- (六) 回答の根拠となった関連文書名を表示する
- (七) 回答内容の信頼度スコアを数値で表示する

解答群

	a	b	c
ア	(一)	(三)	(六)
イ	(一)	(三)	(七)
ウ	(一)	(四)	(六)
エ	(一)	(五)	(六)
オ	(一)	(五)	(七)
カ	(二)	(三)	(六)
キ	(二)	(三)	(七)
ク	(二)	(四)	(六)
ケ	(二)	(四)	(七)
コ	(二)	(五)	(七)

問4 住宅設備メーカーのA社は、設備の製造、販売、保守を行っている。A社では、様々なデータソースから収集したデータを利活用するためのデータ分析基盤を運用している。今回、設備の利用者（以下、ユーザーという）からの問合せに関するデータ利活用の追加要望があり、データ分析基盤へのデータ追加を検討することにした。

〔データ分析基盤〕

A社のデータ分析基盤の全体像を図1に示す。設備、ユーザーなどのマスターデータはDWH内のテーブルに格納している。

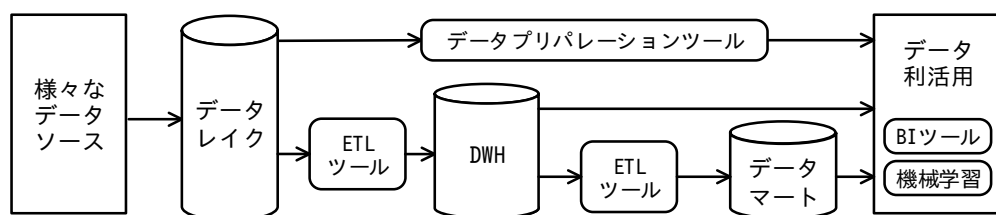


図1 A社のデータ分析基盤の全体像

〔問合せ記録〕

1. コールセンターでは、ユーザーから電話又はWebで受けた設備使用上の不具合や違和感などの問合せを記録している。
2. 問合せ記録は、日別にJSON形式のテキストファイルに保存する。
3. 問合せ記録には、一意な問合せ番号、受付日時、問合せ内容、ユーザーID、ユーザーの連絡先、対象設備、点検修理要否などに加えて、通話又はメールによる発信ごとの発信日時、発信内容が含まれる。

〔データ利活用の要件〕

問合せ記録のデータ利活用要件を表1に示す。

表1 問合せ記録のデータ利活用要件

要件	内容
要件1	設備ごとの問合せ傾向、修理への転換率、問題解決時間などの分析に全社で利用したい。利用にあたっては、蓄積した全期間の問合せを対象に、ユーザーの属性（居住地域、家族構成など）、設備の属性（モデル、容量など）を軸に汎用的に検索、集計を行って可視化することもある。特定の設備を対象に、問合せ内容を照会する場合もある。
要件2	設備保守部門で、一定期間に点検修理を要した問合せを対象に、対象の設備情報（設置場所、設置年月など）、設備利用年月数の導出値を切り出して分析に使用したい。
要件3	受付日時、対象設備、問合せ内容を機械学習の対象にすることで、センチメント分析、FAQの改善を行いたい。

〔データ配置の整理〕

データエンジニアのBさんは、問合せ記録のデータ利活用要件を基に、データ分析基盤へのデータ配置を次のように整理した。

1. 要件1について、問合せ記録の詳細に、転換率、問題解決時間などを付加し、軸となるユーザー、設備のデータを に配置し、 のデータを ETL ツールで加工して格納する。
2. 要件2について、分析に必要なデータを に配置し、 のデータを ETL ツールで加工して格納する。
3. 要件3について、データレイクに格納したデータをデータプリパレーションツールで加工して利用する。

設問 本文中の a ～ d に入れる字句の組合せはどれか。解答群の中から最も適切なものを選べ。

解答群

	a	b	c	d
ア	DWH	データレイク	DWH	データレイク
イ	DWH	データレイク	データマート	データレイク
ウ	DWH	データレイク	データマート	DWH
エ	データマート	DWH	データマート	DWH
オ	データマート	DWH	DWH	データレイク
カ	データマート	データレイク	DWH	データレイク
キ	データマート	データレイク	データマート	DWH
ク	データマート	データレイク	データマート	データレイク

問5 A社は、不動産賃貸仲介業を全国規模で行っている。RDBMSを用いて物件情報検索システム（以下、検索システムという）を運用している。

検索システムは、物件を管理するシステムを補完するシステムであり、担当者が接客するとき、当該システムの“物件”テーブルを利用している。

1. 担当者による接客業務の概要

- (1) 物件を探している借主に対して、当該借主の希望に近い物件を探す支援を行い、借主と貸主との間の交渉・賃貸契約の仲介を行う。
- (2) 物件の貸主に対して、物件の審査を行う。当該貸主に長期の空き物件がある場合、周辺の競合物件の付帯設備（以下、設備という）の設置状況を調査し、当該空き物件に人気の設備を增強することなど、物件の付加価値を高める対策の助言を行うこともある。

2. “物件”テーブル

- (1) A社が仲介する全ての物件を、物件コードで一意に識別する。
- (2) 物件の沿線、最寄駅、賃料、間取りなどの基本属性を記録する列がある。
- (3) エアコン、オートロックなどの設備が設置されているかどうかの有無を記録する列があり、一つの物件に20個の設備の有無を記録できる。
- (4) “物件”テーブルの全ての列にNOT NULL制約を指定している。

3. “物件”テーブルのテーブル構造、主な列の意味と制約及び主な統計情報

“物件”テーブルのテーブル構造を図1に、主な列の意味・制約を表1に、RDBMSの機能を用いて取得した主な統計情報を表2に示す。

物件（物件コード、物件名、沿線、最寄駅、賃料、間取り、向き、専有面積、築年数、都道府県、市区町村、エアコン、オートロック、…）

図1 “物件”テーブルのテーブル構造（一部省略）

表1 “物件”テーブルの主な列の意味・制約（一部省略）

列名	意味・制約
物件コード	物件を一意に識別するコード
沿線	物件から利用可能な最寄駅のうち代表的な沿線の名前
エアコン、オートロック、…	当該設備が設置されているかどうかの有無を示す値 Y：設置あり、N：設置なし

表2 “物件” テーブルの主な統計情報

テーブル名	行数	列名	列値個数
物件	1,600,000	物件コード	1,600,000
		沿線	400
		エアコン	2
		オートロック	2

運用部門では検索システムの処理を把握するため、現在検索できる設備の組合せを述語に指定した SQL 文を調査した。その SQL 文の例を、表3に示す。また SQL 文の結果行を保存するファイルの所要量を見積もる目的で、表3の各 SQL 文の結果行数を見積もった。見積りに当たっては、沿線、エアコン、オートロックの列値の分布は互いに独立であり、各列の列値は一様分布に従うと仮定した。

表3 設備の組合せを述語に指定した SQL 文の例（未完成）

SQL	SQL 文の構文（上段：目的，下段：構文）	見積もった結果行数
SQL1	沿線が○△線であり、かつ、設備にエアコンとオートロックの両方がある物件を調べる。	1,000
	SELECT 物件コード, 物件名 FROM 物件 WHERE 沿線 = '○△線' AND (エアコン = 'Y' AND オートロック = 'Y')	
SQL2	沿線が○△線であり、かつ、設備にエアコン又はオートロックのいずれかがある物件を調べる。	a
	SELECT 物件コード, 物件名 FROM 物件 WHERE 沿線 = '○△線' AND (エアコン = 'Y' OR オートロック = 'Y')	
SQL3	設備にエアコンとオートロックの両方がある物件を沿線ごとに集計した物件数が、全物件数に占める割合を百分率（小数点以下切捨て）で求める。	400
	WITH TEMP (TOTAL) AS (SELECT COUNT(*) FROM 物件) SELECT 沿線, FLOOR(b) FROM 物件 CROSS JOIN TEMP WHERE エアコン = 'Y' AND オートロック = 'Y' GROUP BY 沿線, TOTAL	

注記 FLOOR 関数は、引数以下の最大の整数を計算する。

設問 表 3 中の a , b に入れる字句の組合せを，解答群の中から選べ。

解答群

	a	b
ア	1,000	COUNT(*) / TOTAL
イ	1,000	COUNT(*) * 100 / TOTAL
ウ	1,000	SUM(*) / TOTAL
エ	3,000	COUNT(*) / TOTAL
オ	3,000	COUNT(*) * 100 / TOTAL
カ	3,000	SUM(*) / TOTAL
キ	4,000	COUNT(*) / TOTAL
ク	4,000	COUNT(*) * 100 / TOTAL
ケ	4,000	SUM(*) / TOTAL

問6 次の記述中の a に入れる正しい答えを、解答群の中から選べ。ここで、配列の要素番号は1から始まる。

関数 correction は、100 点満点の試験の得点を昇順に格納した整数型の配列を引数 score で受け取り、ある方法で補正した得点を配列に格納して返す。ただし、受け取る得点は 0 以上の整数とし、0 又は 100 が過半数を占めることはないものとする。

関数 correction を correction({20, 30, 30, 30, 60, 80}) として呼び出したときの戻り値の配列には、値が 50 以上の要素が a 個含まれる。

[プログラム]

```
○整数型の配列: correction(整数型の配列: score)
  整数型: n ← scoreの要素数
  整数型: i, m
  実数型: center // scoreの中央値
  整数型の配列: cvalue ← {n個の 未定義の値} // 補正後の得点
  実数型: buff

  if (n が奇数)
    m ← (n + 1) ÷ 2 の商
    center ← score[m]
  else
    m ← n ÷ 2 の商
    center ← (score[m] + score[m + 1]) ÷ 2 // 実数として計算する
  endif

  for (i を 1 から n まで 1 ずつ増やす)
    if (score[i] が center 以下)
      buff ← (50 ÷ center) × score[i]
    else
      buff ← (50 ÷ (100 - center)) × (score[i] - center) + 50
    endif
    cvalue[i] ← buff を小数第1位で四捨五入した値
  endfor
  return cvalue
```

解答群

ア 1	イ 2	ウ 3
エ 4	オ 5	カ 6

プロフェッショナルデジタルスキル（データ・AI）試験（仮称）

科目 B のサンプル問題 解答例

問番号	正解
問 1	イ
問 2	キ
問 3	エ
問 4	ウ
問 5	オ
問 6	オ

出題趣旨

組織が事業の戦略立案や業務改善を行う上では、事業及び業務のデータを活用したデータドリブンな意思決定と、それを支えるデータマネジメント、データガバナンス、データ基盤の構築が重要である。また、AI を活用する上でも、AI が生成するデータを多角的に分析し、誤りや偏りなく評価した上で事業及び業務に用いる能力が強く求められている。

本サンプル問題では、データ及び AI を活用する現場で直面する様々な状況や課題を題材とし、データライフサイクル全体を通してデータ及び AI の効果的な活用を推進する専門的な技能が、目標とする水準に達していることを評価する。

問 1 は、データの提供者と利用者が遵守すべきガイドラインの策定を題材としている。データの意味の変更が与える影響、データコントラクトを用いた品質担保とトラスト管理の在り方、提供者と利用者双方の責任を理解し、それらを踏まえたルールを策定する能力を評価する。

問 2 は、EC サイトの利益最大化に向けたデータ分析を題材としている。散布図行列を用いたデータ関係性の可視化や、重回帰分析のモデル評価・解釈を通じて、事業の意思決定につながる洞察をデータから得る能力を評価する。

問 3 は、大規模言語モデル（LLM）を活用した支援システムの構築を題材としている。情報の鮮度や正確性が求められる要件を満たすために、文脈内学習（In-Context Learning）などの技術を用い、信頼性の高い AI 活用を設計する能力を評価する。

問 4 は、データ分析基盤の構築を題材としている。データレイク、データウェアハウス、データマートといった各構成要素の役割と、データの加工・保存の方法に関するデータエンジニアリングの能力を評価する。

問 5 は、不動産業の物件情報検索システムを題材としている。RDBMS における SQL を用いたデータ操作の理解と、統計情報に基づく処理結果（データ量）の予測など、業務に沿ってデータを処理する能力を評価する。

問 6 は、データを補正するアルゴリズムを題材としている。データを活用する際には、データの前処理を行うツールなどを AI も駆使して作成することがある。中央値の算出や条件分岐を含む擬似言語プログラムを読み解き、処理の内容を理解した上で結果を導き出す能力を評価する。