



デジタル時代の人材政策に関する検討会用資料

論点 2.

生成AIがDX推進を担う専門レベルの人材に与える影響
～ データサイエンティスト ～

2023年12月22日

一般社団法人 データサイエンティスト協会 佐伯諭

データサイエンス領域の切り口から3つのことを議論させてください

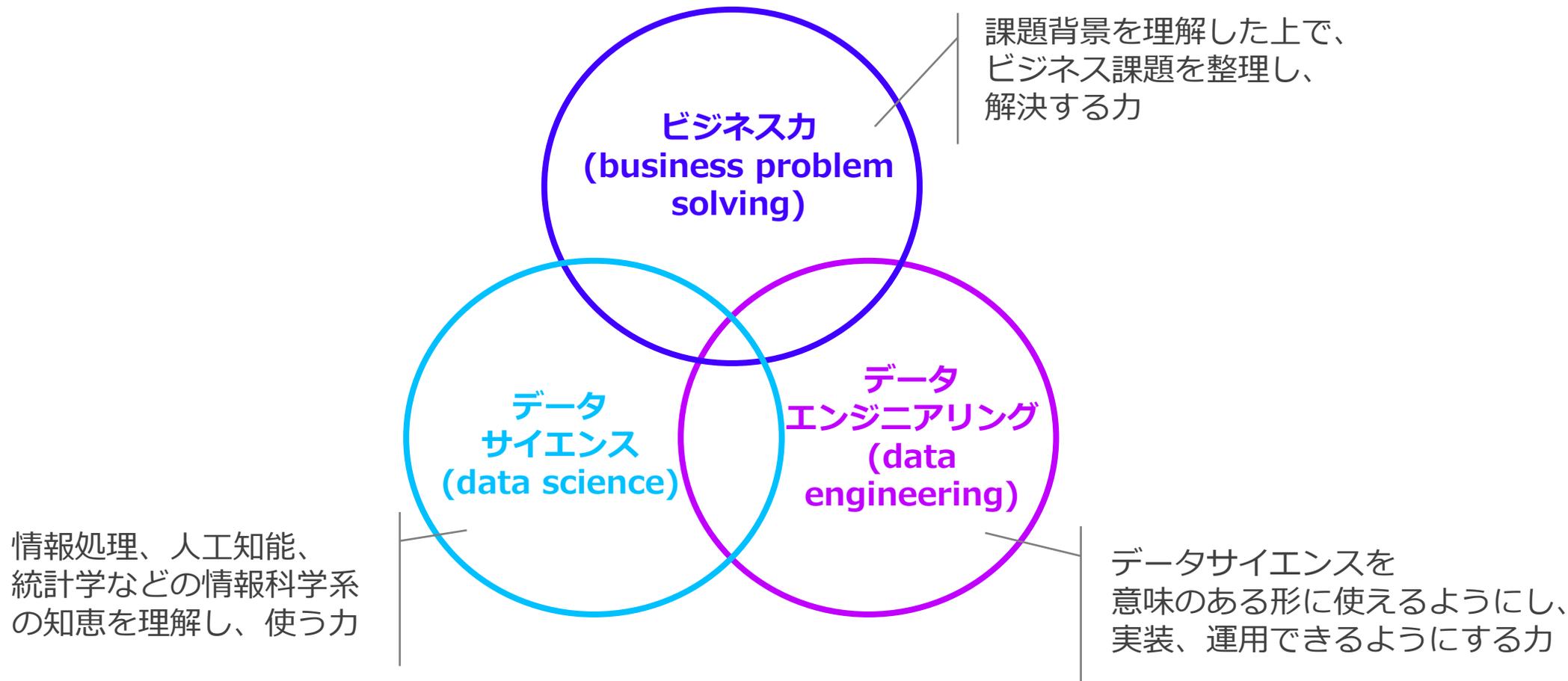
1. DX推進人材自身が生成AIによってどう変わるか？
2. “生成AIを活用した企業DXプロジェクト”におけるDX推進人材の業務や意識
3. 企業間連携や産業を跨いだDX推進に必要な考え方

本資料はDS協会スキル定義委員会での生成AIに関する今年10月の発表内容を元に構成しています

データサイエンス領域の切り口から3つのことを議論させてください

1. **DX推進人材自身が生成AIによってどう変わるか？**
2. “生成AIを活用した企業DXプロジェクト”におけるDX推進人材の業務や意識
3. 企業間連携や産業を跨いだDX推進に必要な考え方

3つのスキルセット



資料 : データサイエンティスト協会プレスリリース (2014.12.10) <http://www.datascientist.or.jp/news/2014/pdf/1210.pdf>

生成AIはDSの業務効率化も実現するが、DSの本質は変わらない

生成AIによって、データサイエンティスト（DS）の業務の一部は効率化する。
だが、「データから価値を創出し、ビジネス課題に答えを出す」というDSの存在価値の本質は変わらない



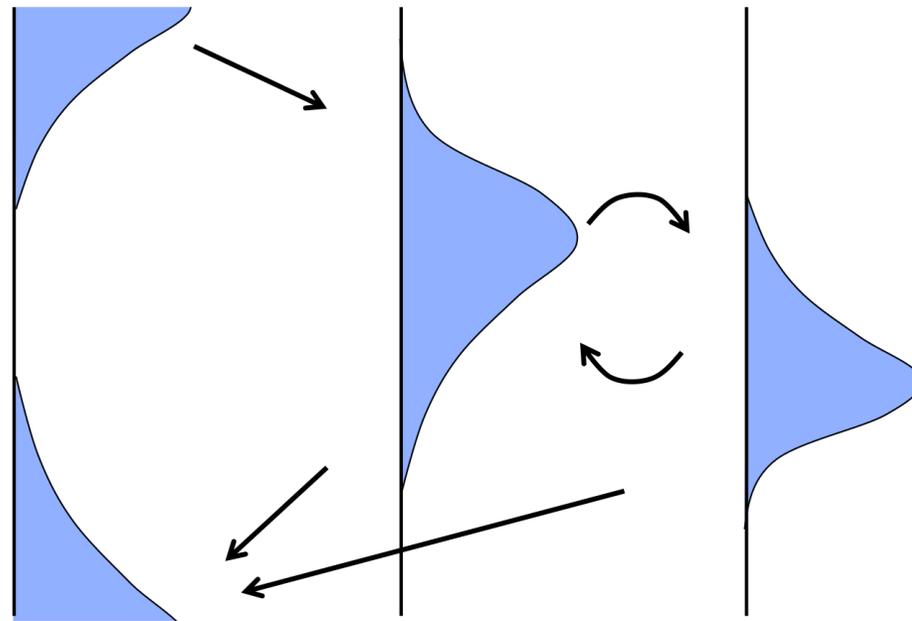
目的・テーマ設定

問題定義

アプローチの設計

処理・分析

解決



資料：データサイエンティスト協会プレスリリース (2014.12.10) <http://www.datascientist.or.jp/news/2014/pdf/1210.pdf>

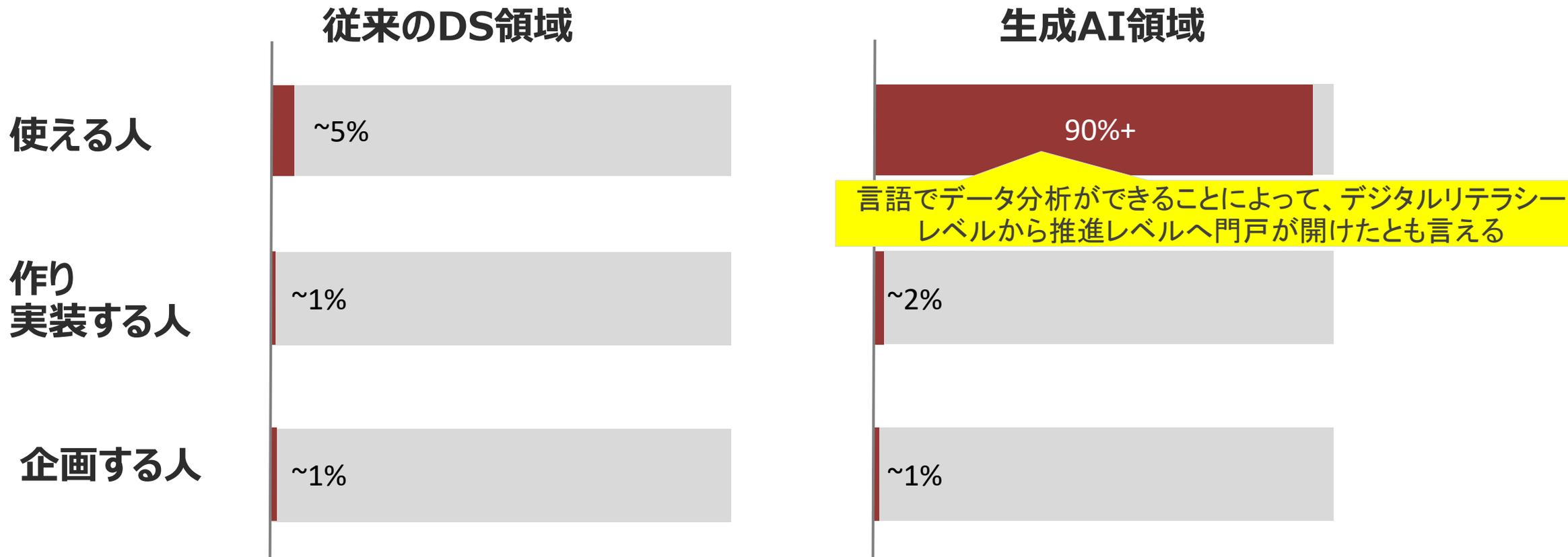
生成AIの活用場面（例）

- ✓ 課題解決アイデアの壁打ち
- ✓ 要約、要点抽出
- ✓ ダミーデータの作成
- ✓ オープンデータ収集
- ✓ Pythonコード生成
- ✓ Advanced Data Analysisでデータ分析やグラフ作成
- ✓ PPT資料作成

使える人は爆増するが、企画や実装する人は高い専門性が必要

数値はイメージ

ChatGPTなどを利用するだけなら“言語を知っている”レベルで活用可能。一方で実装側のハードルは高い。



資料：DS協会 スキル定義委員会討議

“使う”レベルのスキルはプロンプトエンジニアリングの域で十分

スキルレベル	チェック項目
★	LLMを利用して、データ分析やサービス、システム開発のためのコードを作成、修正、改良できる
★	LLMを利用して、開発した機能のテストや分析検証用のダミーデータを生成できる
★	画像生成AIに組み込まれた標準機能の利用（モデル選択）や、画像生成プロンプトルール（強調やネガティブプロンプトなど）を理解し、適切に入力することで、意図した画像を生成できる
★	既存の生成AIサービスやツールを活用し、自身の身の回りの業務・作業の効率化ができる
★	生成AIを活用する際、出力したい要件に合わせ、Few-shot PromptingやChain-of-Thoughtなどのプロンプト技法の利用や、各種APIパラメーター（Temperatureなど）の設定ができる
★★	大規模言語モデルでの Hallucination に惑わされないために、どのような質問は適切で、どのような質問が適切でないかを判断して利用できる
★★	独自に用意したプロンプト入力用のデータベースやファイルを用いて、プロンプトに追加することで、意図した文章を生成することができる
★★	画像生成AIに組み込まれた標準機能の利用（モデル選択）や、画像生成プロンプトルール（強調やネガティブプロンプトなど）を組み合わせ、入力を試行錯誤することで、より意図に即した画像を生成できる

資料：DS協会 スキルチェックリストVer.5より抜粋

2つのDSスキル

主たる情報科学技術

- 統計数理 (Stats)
- 古典的機械学習 (ML)
- 深層学習 (DL)
- 敵対的生成ネットワーク (GAN)
- 拡散モデル (diffusion model)
- 大規模言語モデル (LLM)

先端現場での導入

1975-

2005-

2014-

2018-

2020-

2022-



スキルのイメージ

いわゆる
Data Analyst
DataScientist
スキル

AIを使い倒す
スキル (AI利
活用スキル)

資料: データサイエンティスト協会 スキル定義委員会 討議 (2023年6月)

データサイエンス領域の切り口から3つのことを議論させてください

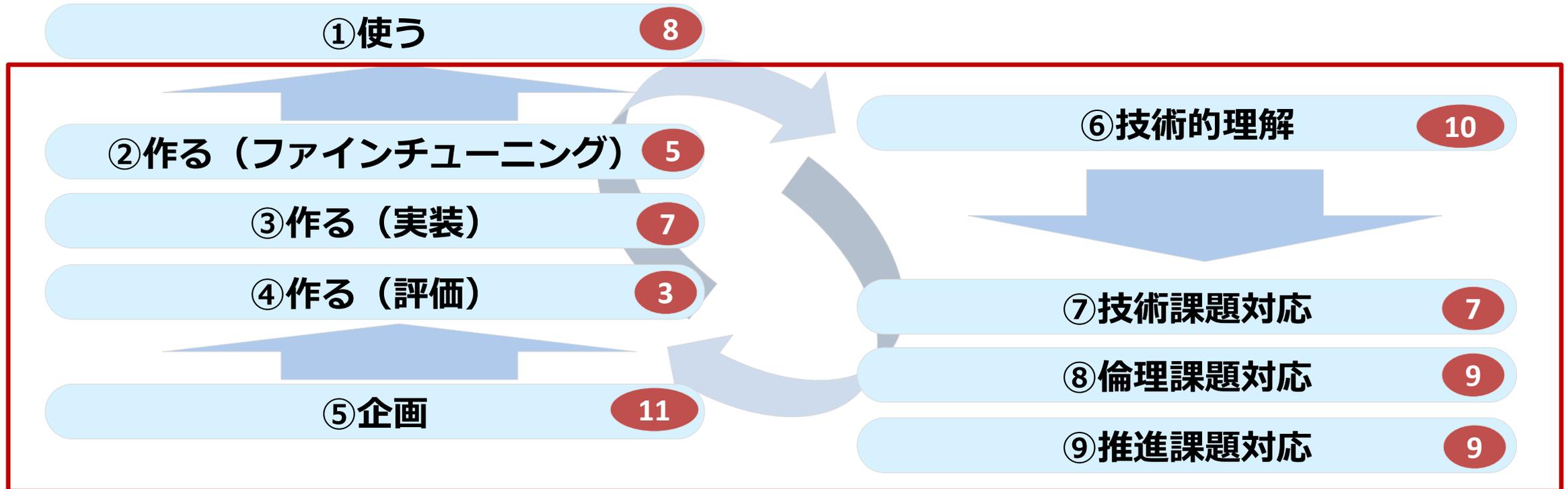
1. DX推進人材自身が生成AIによってどう変わるか？
2. “生成AIを活用した企業DXプロジェクト”におけるDX推進人材の業務や意識
3. 企業間連携や産業を跨いだDX推進に必要な考え方

AI利活用スキルは大きく2つに類型化される

生成AIは利用・開発・企画といった活用するためのスキルが当然必要となる一方で、その技術的背景を理解しつつ、生じる課題に対応するスキルが求められる

利活用スキル 34

背景理解・対応スキル 35



資料: データサイエンティスト協会 スキル定義委員会 発表資料 (2023年10月)

AI利活用スキル例

利活用スキル 34

①使う

8

②作る（ファインチューニング）

5

③作る（実装）

7

④作る（評価）

3

⑤企画

11

★ ★ ファインチューニングの目的や要件に対して、ファインチューニング用のデータ収集、必要に応じたデータベース化を行い、ファインチューニングを実施・評価できる（ベクトルDB、LlamaIndex、LangChain、モデレーションなど）

★ ★ 拡散モデルの利用要件を踏まえ、LoRAやControlNetなどを用いて、ファインチューニングできる

★ ★ 生成AIが提供するAPIやプラグインなどの機能を用いてシステム連携することで、自社の新機能やサービスを設計・実装できる

★ ★ 基準となる指標や評価のためのデータセットなどの性能評価方針に従い、大規模言語モデルの性能を評価できる

AI利活用スキル例

利活用スキル 34

①使う

8

②作る（ファインチューニング）

5

③作る（実装）

7

④作る（評価）

3

⑤企画

11

★★ 低品質なデータ（バイアスや欠落、計測・登録の誤り、混入など）に対するデータクレンジングには限界があり、取得段階での対策が重要であることを理解している

★★ 用途・目的に応じ、生成AI（LLMやDiffusionモデルなど）の選択、開発と運用時両面の計算リソース、著作権侵害リスクや商用利用可否などを総合的に判断し、活用の検討ができる

★★★ 新たなテクノロジー・デバイスやAIサービスなどが登場した際に、速やかにそれらを活用・応用した新たなサービスの企画・設計や、データ活用戦略が立案できる

資料：データサイエンティスト協会 スキル定義委員会 発表資料（2023年10月）

AI利活用スキル例



様々なサービスが登場する中で直感的にわくわくし、その裏にある技術に興味を持ち、リサーチできる



拡散モデルや大規模言語モデルなど生成モデルに関する技術の動向や広がり理解し、どのような生成タスク分野に適用可能かを調査・検討できる
(テキスト生成、画像生成、音源生成、ダミーデータ生成など)



過去に何度か起こった技術的ブレイクスルーとその後の世界の変化のあり方を理解した上で（深層学習や深層強化学習、GAN）、今起こっているブレイクスルー（拡散モデルや大規模言語モデル等）の影響や起こすべき行動の見立てを立てられる

背景理解・対応スキル 35

⑥技術的理解 10



⑦技術課題対応 7

⑧倫理課題対応 9

⑨推進課題対応 9

資料: データサイエンティスト協会 スキル定義委員会 発表資料 (2023年10月)

AI利活用スキル例

★★

大規模言語モデル（LLM）でHallucinationが起こる理由を学習に使われているデータの観点から説明できる（学習用データが誤りや歪みを含んでいる場合や、入力された問いに対応する学習用データが存在しない場合など）

★★

生成AIモデルを活用する際、倫理の曖昧さと法整備の不十分さが残っていることを理解し、利活用するための対応方法を検討できる

★★★

プロジェクトやチームを超えて、データサイエンス・AIに関する正しい理解、知識・スキル向上のための取り組みを企画・推進できる

★★★

マルチモーダルAIの今後の発展を踏まえ、自社の持つインフラ環境の限界を考慮し、データ活用戦略の見直しやデータ基盤の再構築を検討できる

背景理解・対応スキル 35

⑥技術的理解

10

⑦技術課題対応

7

⑧倫理課題対応

9

⑨推進課題対応

9

資料：データサイエンティスト協会 スキル定義委員会 発表資料（2023年10月）

AI利活用スキルのスキルカテゴリ・スキルレベルの広がり

スキルカテゴリ

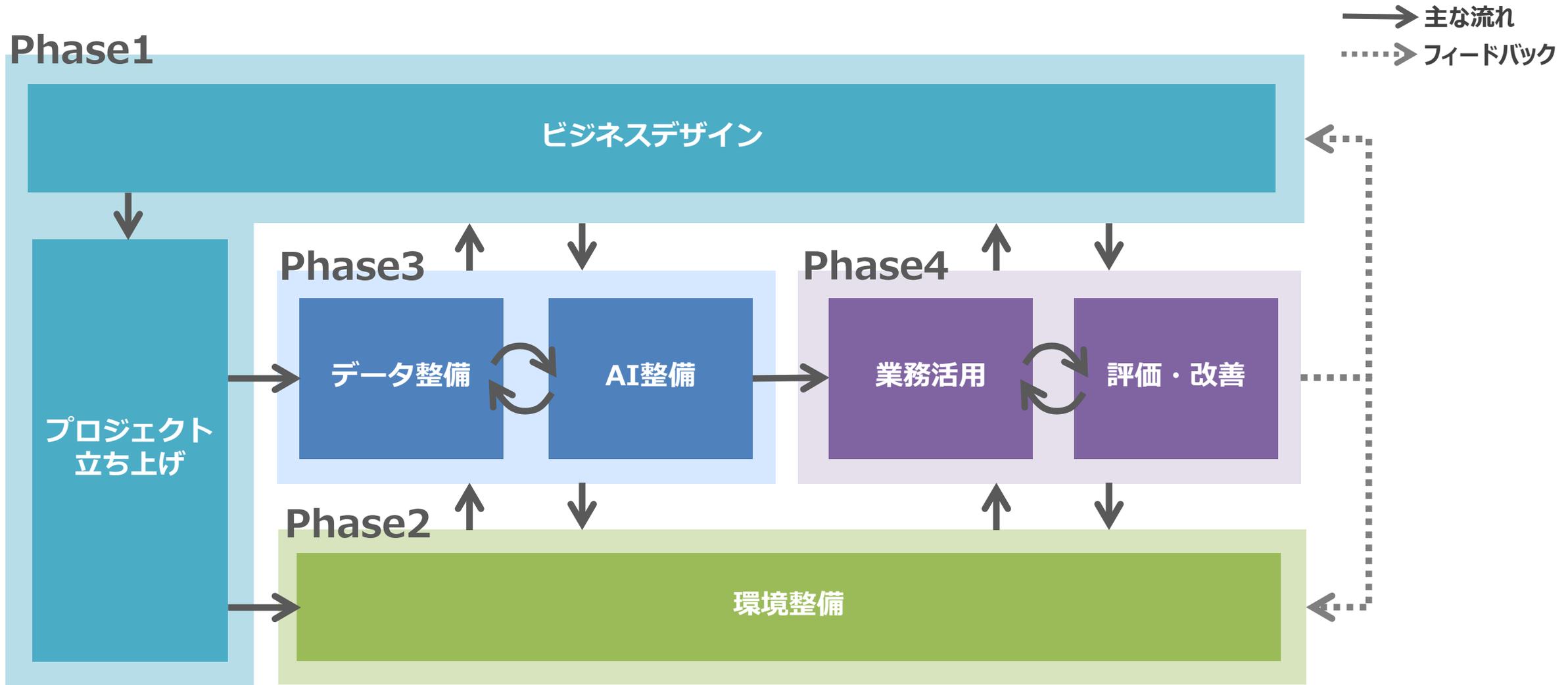
フェーズ	ビジネス	データサイエンス	データエンジニアリング
①使う	3		5
②作る-FT		2	3
③作る-実装	2	1	4
④作る-評価		3	
⑤企画	8	3	
⑥技術的理解	4	6	
⑦技術課題対応	4	3	
⑧倫理課題対応	9		
⑨推進課題対応	8	1	
総計	38	19	12

レベル

フェーズ	★	★★	★★★
①使う	5	3	
②作る-FT		2	3
③作る-実装		1	6
④作る-評価		1	2
⑤企画	1	7	3
⑥技術的理解	1	4	7
⑦技術課題対応	3	1	3
⑧倫理課題対応	3	4	2
⑨推進課題対応		2	7
総計	13	25	31

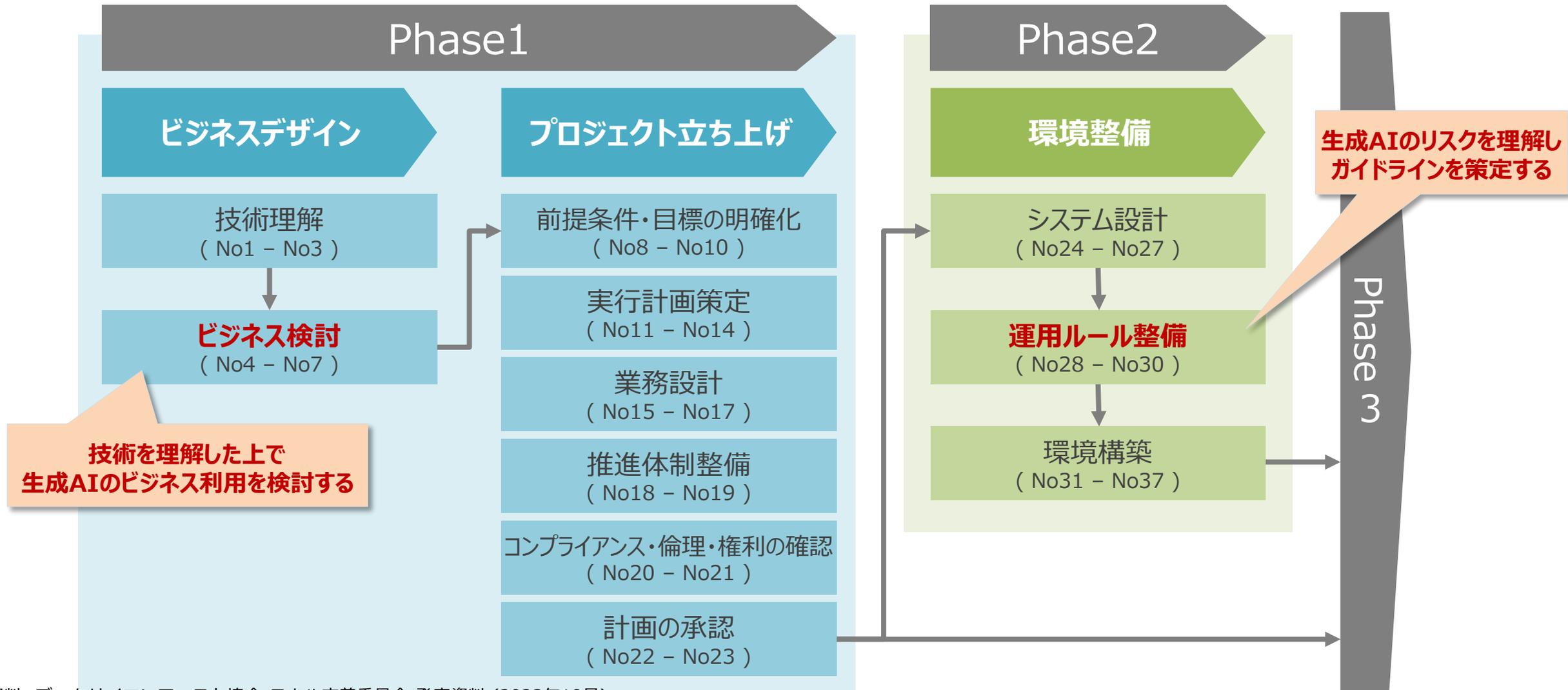
資料: データサイエンティスト協会 スキル定義委員会 発表資料 (2023年10月)

AI活用タスクリスト<構造図>



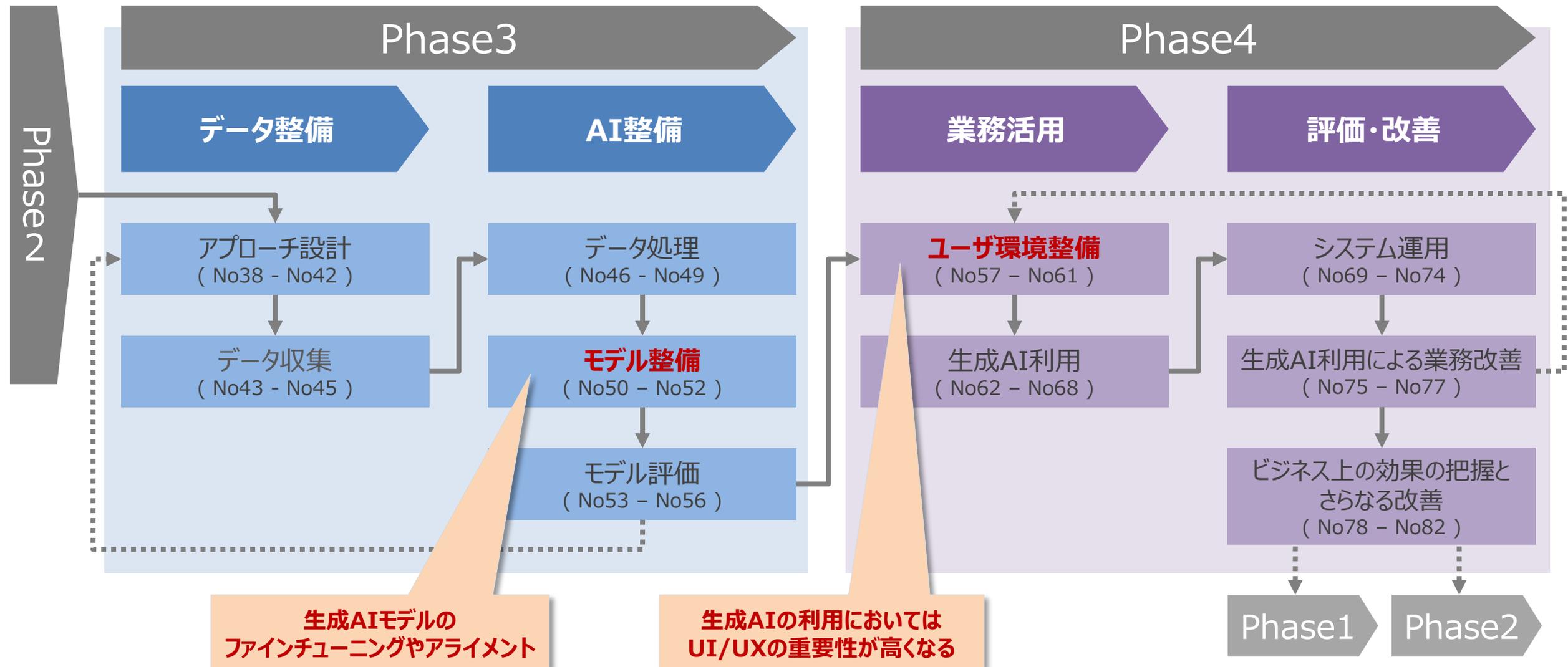
資料: データサイエンティスト協会 スキル定義委員会 発表資料 (2023年10月)

AI利活用タスクリスト <Phase1~2>



資料: データサイエンティスト協会 スキル定義委員会 発表資料 (2023年10月)

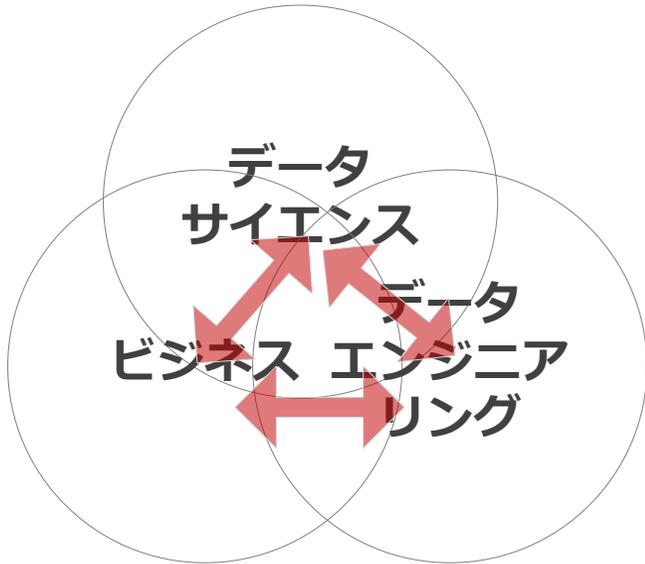
AI利活用タスクリスト<Phase3~4>



資料: データサイエンティスト協会 スキル定義委員会 発表資料 (2023年10月)

AI利活用スキルはそれぞれの分野スキルの融合度が高い

AI利活用プロジェクト

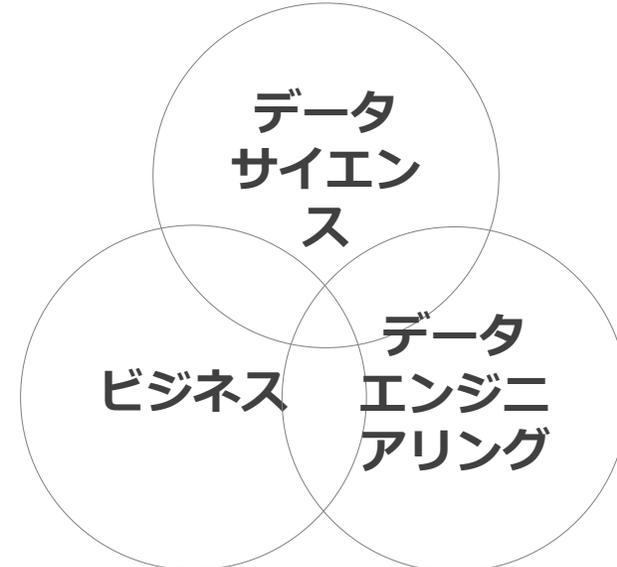


プロジェクト推進役の棟梁クラスは激しく交差したスキルが必要

★★★ スキル例

新たなテクノロジー・デバイスやAIサービスなどが登場した際に、速やかにそれらを活用・応用した新たなサービスの企画・設計や、データ活用戦略が立案できる

近年のデータサイエンス関連プロジェクト



それぞれに高度な専門性が必要
大きなプロジェクトはチームで対応

DS：予測モデル構築において頑健性（Robustness）を維持するための具体的な方法を設計、実施できる

DE：扱うデータのデータ規模や機密性、分析要件を理解した上で、オンプレミスで構築するか、クラウド上で構築するかの要件を整理できる

資料：データサイエンティスト協会 スキル定義委員会 発表資料（2023年10月）

データサイエンス領域の切り口から3つのことを議論させてください

1. DX推進人材自身が生成AIによってどう変わるか？
2. “生成AIを活用した企業DXプロジェクト”におけるDX推進人材の業務や意識
3. **企業間連携や産業を跨いだDX推進に必要な考え方**

データ流通を推進することで様々な生成AIサービスが国内で創出される

■ 生成AIプロジェクトを推進する企業によくある課題

- ✓ データが学習に使われないとはいえ、海外クラウド環境に社外秘情報を送ることのリスクへの対応
- ✓ 自社データだけでAI対話サービスを作ってもビジネス初心者の域。誤答も多く、良い性能が出ない
- ✓ 検証、チューニングコストが高い。POC疲れの再来



■ 企業が求める新たなニーズ（誰かがいい感じに作ったチューニング済みモデルを使わせて～）

企業間連携や産業データ全体を活用した特化型国産基盤モデル

- ✓ 企業内データを安全な方法で移動するだけで良い答えが返ってくる基盤モデル
- ✓ 業界団体などでの生成AI利用、クラウド利用ガイドラインの見直し/策定
- ✓ 情報流通インフラ整備
- ✓ 一企業が海外製生成AIではなく、（おそらく費用が高い）国産生成AIを使うことへのインセンティブ設計