



DataScientist Society

10thシンポジウム

2023年度スキル定義委員会活動報告 2023年度版スキルチェック&タスクリスト公開

2023年10月20日

一般社団法人 データサイエンティスト協会 スキル定義委員会

パネリスト紹介

安宅 和人	スキル定義委員会 委員長 慶應義塾大学SFC教授 / LINEヤフー株式会社 シニアストラテジスト
菅 由紀子	株式会社Rejoui 代表取締役
孝忠 大輔	日本電気株式会社 AI・アナリティクス統括部長
佐伯 諭	スキル定義委員会副委員長 データサイエンティスト協会 事務局長
杉山 聡	株式会社アトラエ データサイエンティスト / AIcia Solid Project 主催
高橋 範光	株式会社デジタルグロースアカデミア 代表取締役社長
田中 貴博	株式会社日立アカデミー 研修開発本部部長
森谷 和弘	データ解析設計事務所 代表 / データアナリティクスラボ株式会社CTO

他 スキル定義委員会メンバー

アジェンダ

1. スキル定義委員会の活動
 - スキル定義委員会とは？
 - 主な活動紹介
2. 生成AIとデータサイエンティストの関係考察
3. AI利活用スキル
4. スキルチェックリスト ver.5
5. データサイエンス業務のタスクリスト ver.4（IPAとの協働作業）
6. ディスカッション

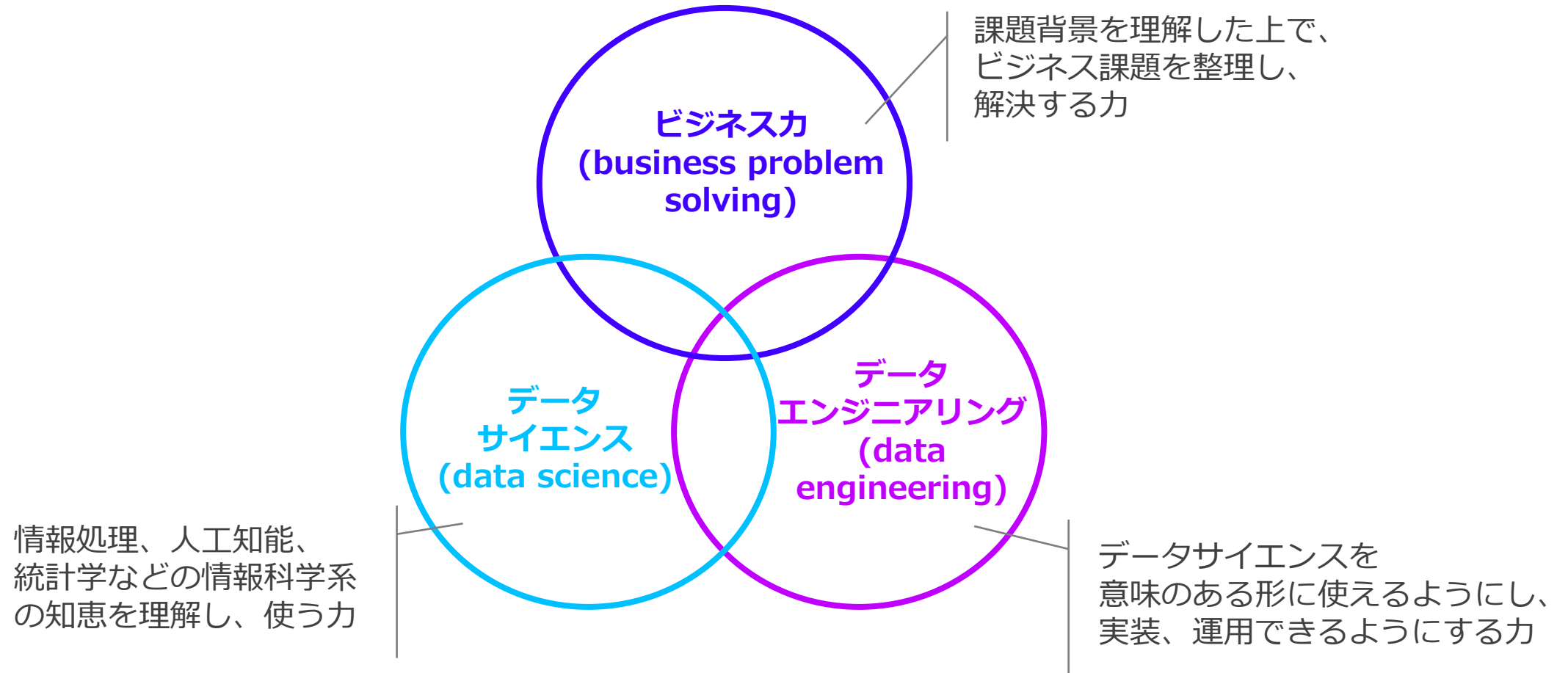
設立時の状況

- ・ バズワードである「データサイエンティスト」という言葉の定義が欠落
- ・ 雇い主側の期待に雇われる側のスキルセットが合致しないケースが増加
- ・ 若い才能たちが、自分達をどう訓練し、スキルを身に付けていくべきかわからない

設立の目的

- ・ この新しいデータプロフェッショナル（「データサイエンティスト」）に必要とされるスキルセットを定義する
- ・ スキル育成と評価のための軸・基準を作る
- ・ このデータプロフェッショナルのみなさんが相互に接し、やりとりできるオープンな環境を提供する
- ・ 社会との対話を行なう

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>

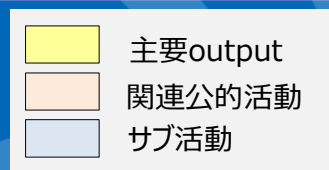
データサイエンス力、
データエンジニアリング力をベースに
データから価値を創出し、
ビジネス課題に答えを出す
プロフェッショナル

スキルレベル

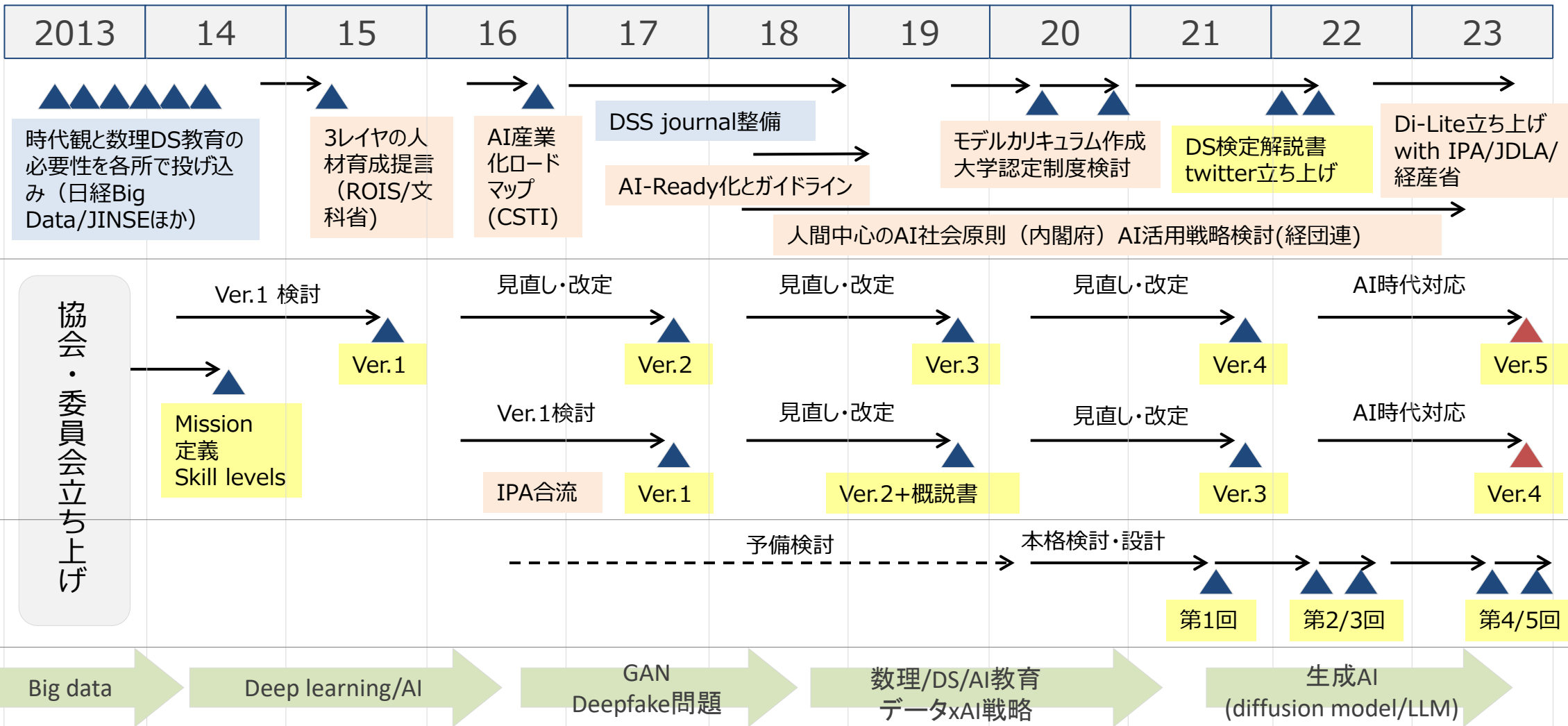


資料：データサイエンティスト協会プレスリリース (2014.12.10) <http://www.datascientist.or.jp/news/2014/pdf/1210.pdf>

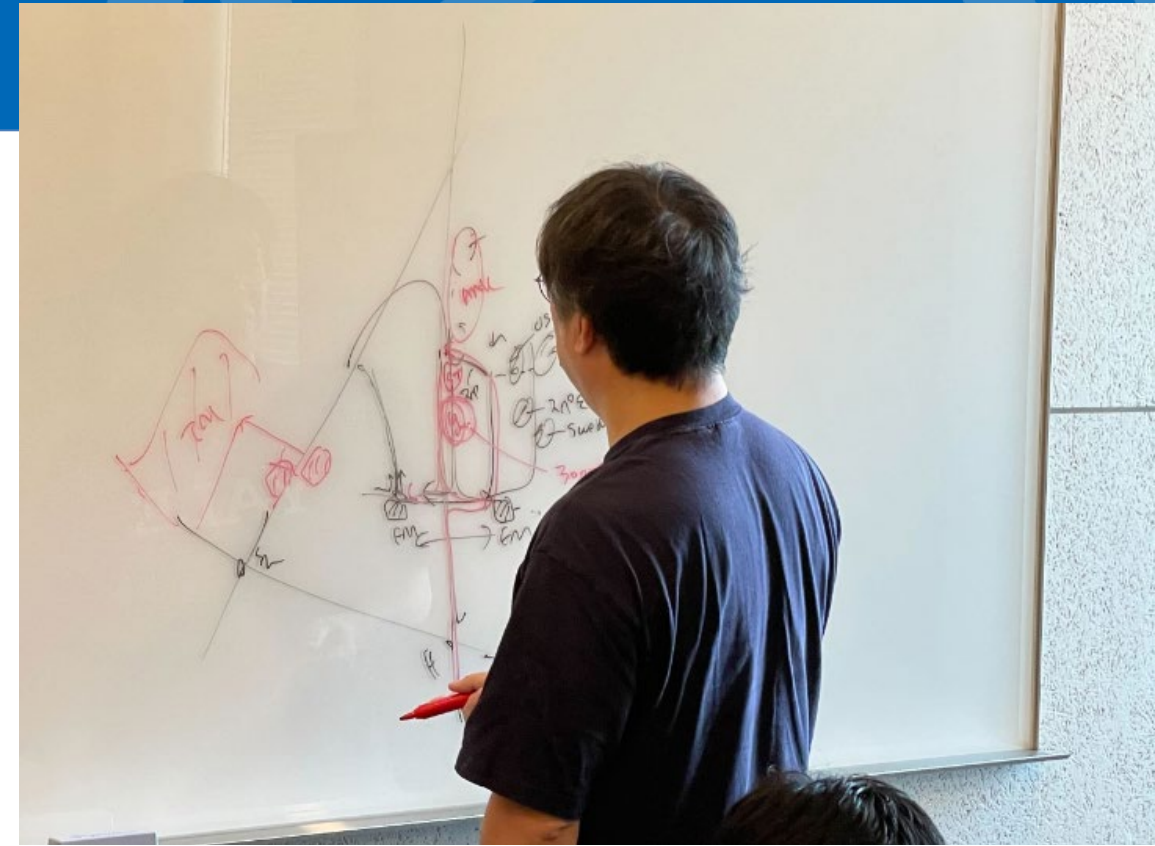
スキル定義委員会のこれまでの取り組み（概要）



協会設立 ▼



毎年恒例の合宿の様子 2023年8月@福岡

[illegible]

生成AI活用アイデアコンテスト！

巨額の賞金つき！ソフトバンクが月間2,500万円のアイデアコンテストを開催！AIを中心とした戦略とは

 神田敏晶  エキスパート | ITジャーナリスト・ソーシャルメディアコンサルタント
6/22(木) 13:46



		生成AI活用コンテスト (グループ企業内で毎月開催へ)
提案件数 (10日間)	5.2万件	
賞金 (毎月)	優勝 1,000万円 総額 2,500万円	

※提案件数 (5.2万件)：2023年5月9日～5月19日(10日間)に提出された提案件数
※コンテストの開催予定は今後変更となる場合があります。

出典：ソフトバンクグループ 定時株主総会資料

<https://news.yahoo.co.jp/expert/articles/24d6c73a5efa3be448d0a34e2eb4fc6baa002fd8>

画像生成AIを高度に使いこなす日本のクリエイターの世界

2023
10/02 【初心者向け】画像生成AI別のプロンプト（呪文）比較表

■ AIイラスト DallE Midjourney NovelAI Stable Diffusion プロンプト 2023年3月15日 2023年10月2日



<https://blogcake.net/prompt-comparison/>



<https://illust.hananosu.net/fantasy-doraque/>

2023年に入って国産LLMの開発が目白押し

「世界トップレベルの大規模言語モデルの開発に着手」 ——産総研らが表明 目指すのは“GPT-3級の日本語 LLM”

<https://www.itmedia.co.jp/news/articles/2310/18/news121.html>

CEATEC 2023

NECが国産生成AIを初めて一般公開、日本語LLMのデモン ストレーションを実施

[https://monoist.itmedia.co.jp/mn/articles/2310/18/news079.h
tml](https://monoist.itmedia.co.jp/mn/articles/2310/18/news079.html)

国産の大規模言語モデル（LLM）の開発を行う 「SB Intuitions株式会社」が本格的に稼働

https://www.softbank.jp/corp/news/press/sbkk/2023/20230804_02/

2023年8月4日
ソフトバンク株式会社

表：国内での日本語 LLM 開発状況

区分	団体名・企業名	備考
大学・ 研究機関	情報通信研究機構	2023 年 7 月に日本語に特化した LLM を開発したと発表
	産業技術総合研究所	ストックマークと共同開発を実施、同社は 2023 年 8 月に LLM を公開
	理化学研究所	東京工業大学、東北大学、富士通と共同で研究開発を実施
	国立情報学研究所	LLM 勉強会を立上げ、2023 年度中の LLM 構築を予定する
	東京大学 松尾研究室	2023 年 8 月に日英 2 ヶ国語対応の LLM を公開
企業	オルツ	2021 年に初期 LLM を公開、アップデート版を 2023 年 2 月に公開、7 月に個別 LLM 受託開発を行うと発表
	rinna	2021 年に日本語特化 LLM を公開、以後手法や規模が異なる複数種類の LLM を発表、2023 年 8 月に法人向けカスタマイズサービス提供を開始
	ABEJA	2023 年 3 月に LLM を商用サービスとして提供開始
	サイバーエージェント	2023 年 5 月に日本語 LLM を公開
	NEC（日本電気）	2023 年 7 月に日本語 LLM を開発したと発表
	ELYZA	2023 年 8 月に LLM を公開、Llama 2 に日本語の追加学習を行った
	LINE/ワークスモバイルジャパン	2023 年 8 月に日本語 LLM を公開
	NTT（日本電信電話）	LLM の独自開発を行い、2023 年度中に公開する計画
	Preferred Networks	商用化を目指す LLM の開発を発表
	ソフトバンク	2023 年 3 月に LLM の開発を行う新会社 SB Intuitions を設立
	日立製作所	制御・運用技術のナレッジを学習した日立独自の LLM を開発する

出所：公開情報をもとに DTFA インスティテュート作成、「企業」区分は大学・研究機関の備考欄掲載企業を除く

<https://faportal.deloitte.jp/institute/report/articles/000862.html>

ビジネスの現場では
生成AIをどのように
活用していますか？

スキル定義委員会での 議論の振り返り

知的生産現場のuser interface (UI) 進化

a perspective

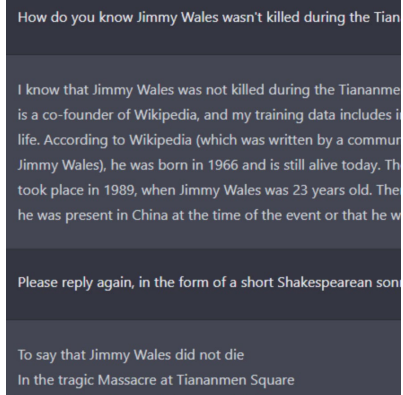
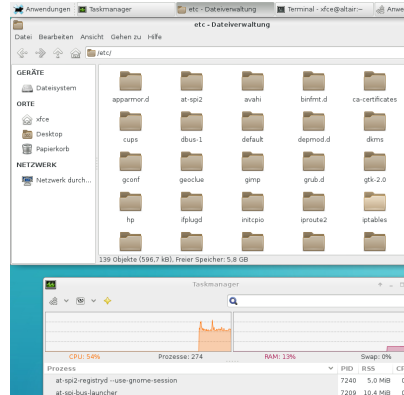
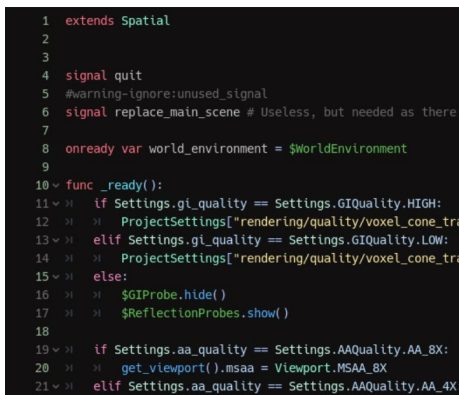
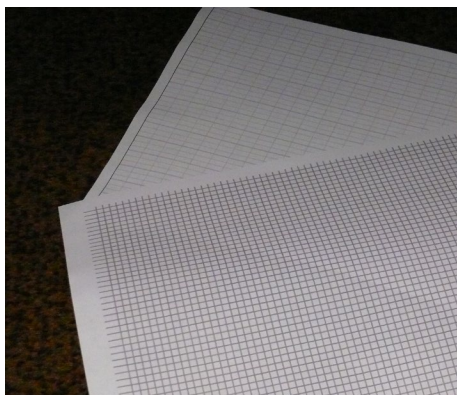
紙

黒画面

Graphical user interface (GUI)

Multi-touch user interface

自然言語



グラフ用紙
原稿用紙

PC
Workstation

Macintosh
Windows 95

iOS (iPhone, iPad)
Android

ChatGPT
Alexa/Google assistant

鉛筆・ペン

プログラミング言語

マウスとキーボード

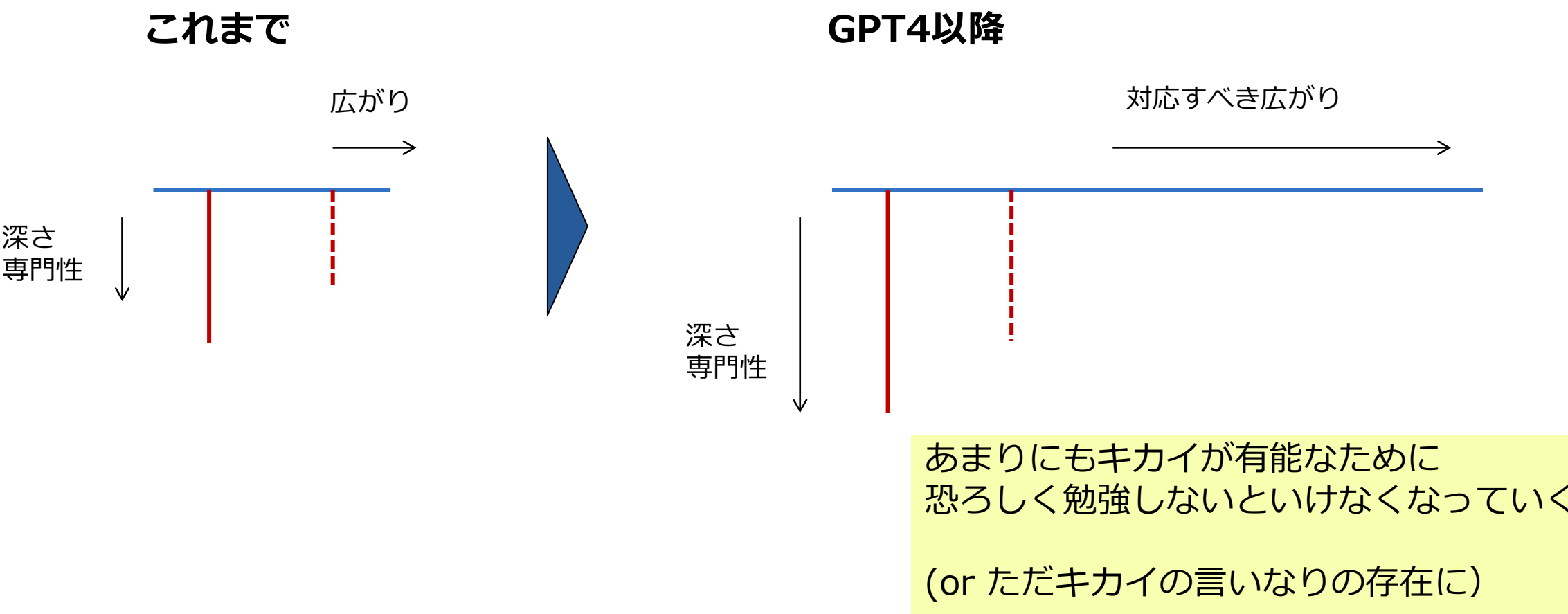
指

テキスト/声

資料：安宅和人「これからの人材育成を考える」文部科学省 今後の教育課程、学習指導及び学習評価等の在り方に関する有識者検討会（第3回） March 24, 2023
<https://commons.wikimedia.org/wiki/File:Gdscript3.4.png> (by SakuraMiyazono, CC BY-SA 4.0) ; <https://commons.wikimedia.org/wiki/File:XFCE-4.12-Desktop-standard.png> (by Martin Wagner, CC BY-SA 3.0) ; https://commons.wikimedia.org/wiki/File:Aldea_Digital_2013_01.jpg (by ProtoplasmaKid, CC BY-SA 3.0) ; https://commons.wikimedia.org/wiki/File:ChatGPT_Wales_Tiananmen_Square_2022-12-30_181035.png (by Rolf h nelson) ; 安宅和人分析

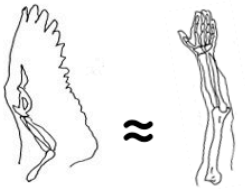
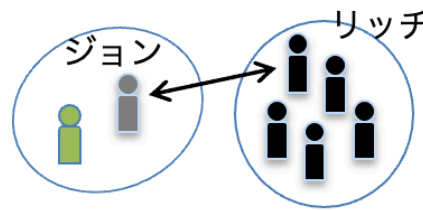
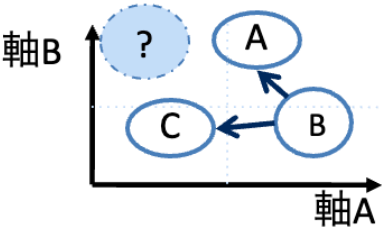
T型、Π型人材の意味の変化

、文理を超えた真に教養の時代に突入



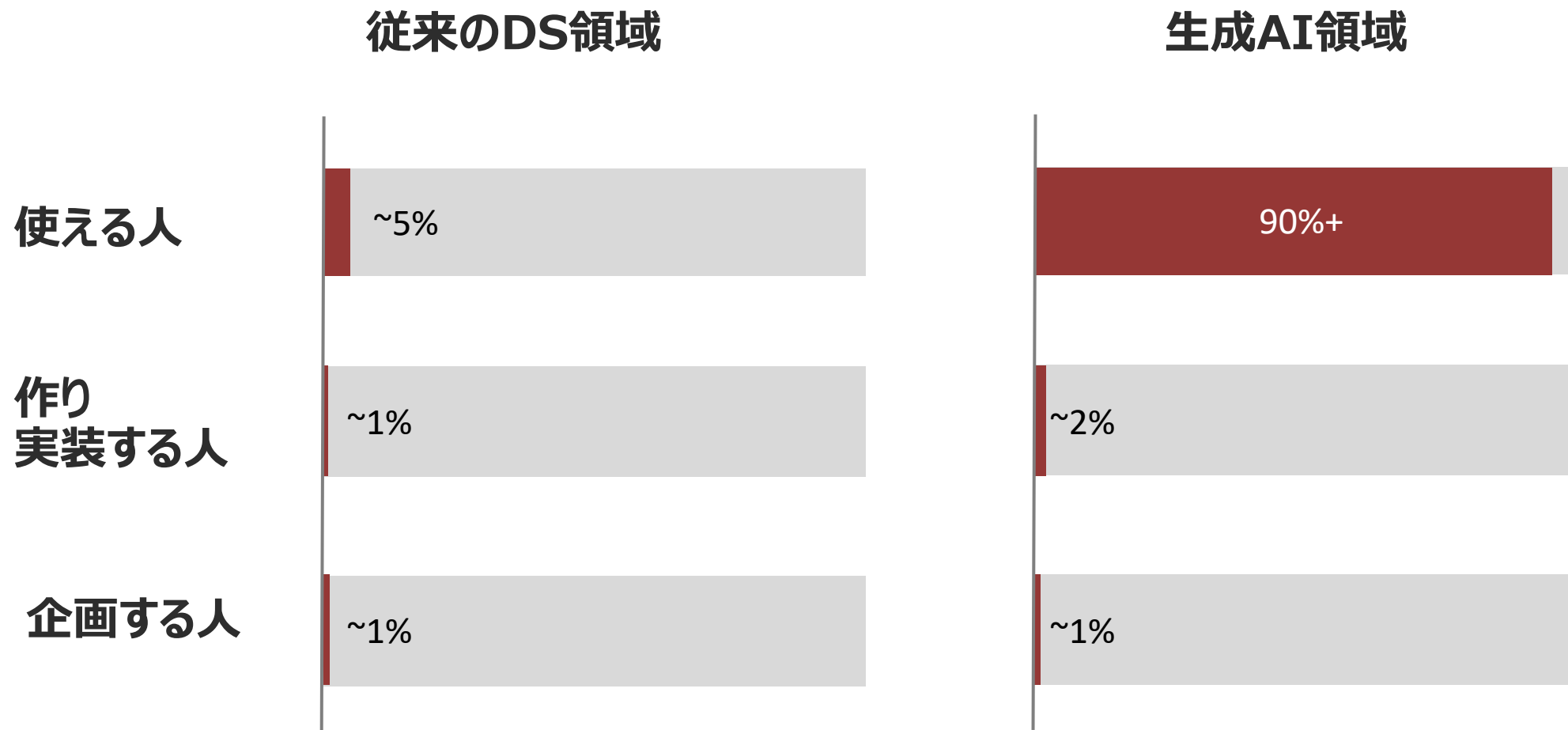
資料： 安宅和人「これからの人材育成を考える」文部科学省 今後の教育課程、学習指導及び学習評価等の在り方に関する有識者検討会（第3回） March 24, 2023
DS協会 スキル定義委員会討議における安宅和人分析 (Mar.21, 2023)

「理解」のベースにあるもの、関係性の発見

	例	正しいときの意味合い
共通性の発見		腕の構造から、翼の構造についての洞察ができると、進化上の比較軸として利用できる
関係性の発見		リッチの言動から、ボールの動きを推測することができる
グルーピングの発見		力学的な関係に加え、これまで気付いていなかった盲点の領域が見えてくる
ルールの発見	$X = \frac{\alpha \times \beta}{\gamma}$	α、β、γの個別要素のそれぞれの変化が全体 (X) にどのように影響するか推測できる

生成AIは使えるかどうかというより、企画し実装できるかどうか

数値はイメージ



資料：DS協会 スキル定義委員会討議

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

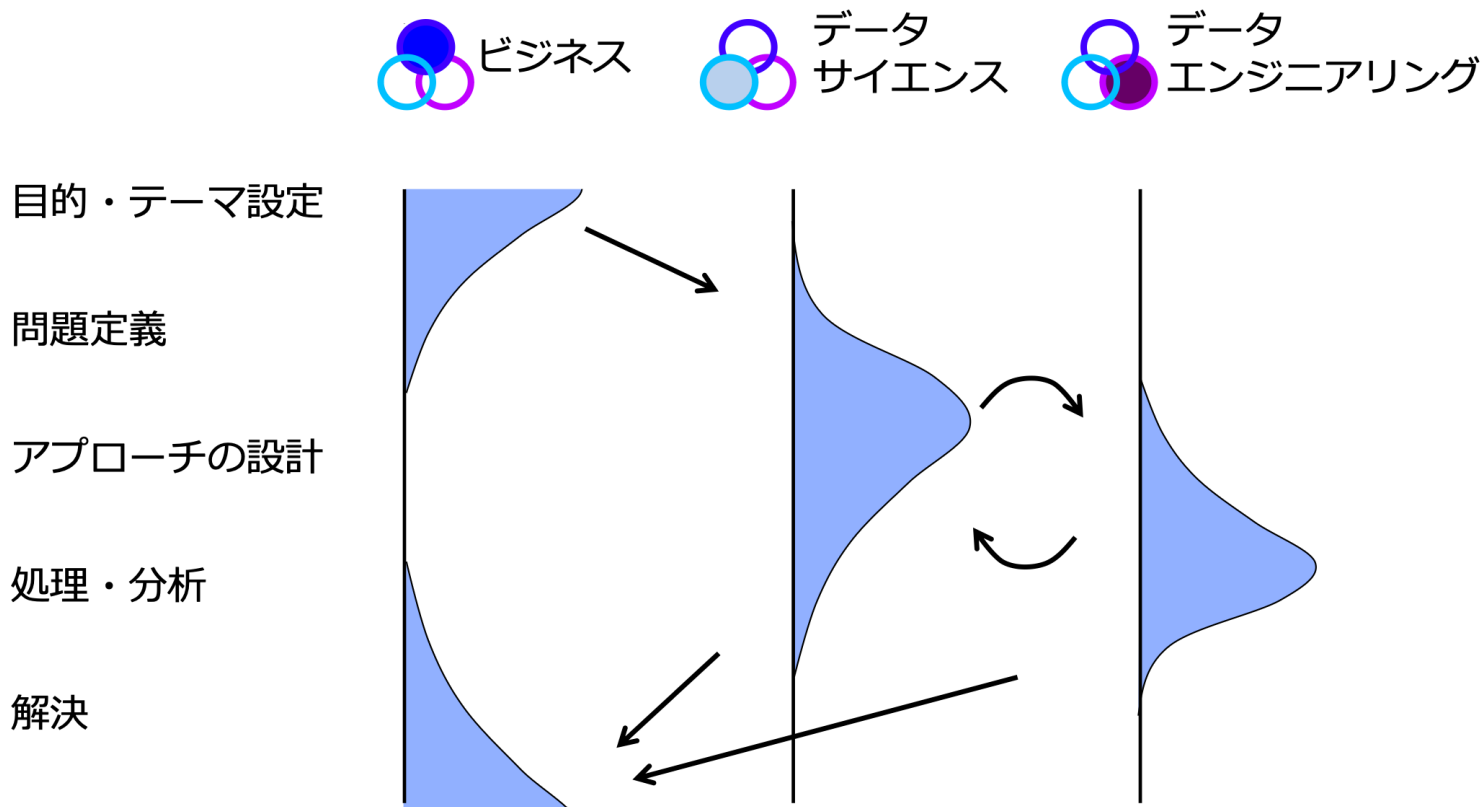
AIを使い倒す
スキル (AI利
活用スキル)

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)

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、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



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）

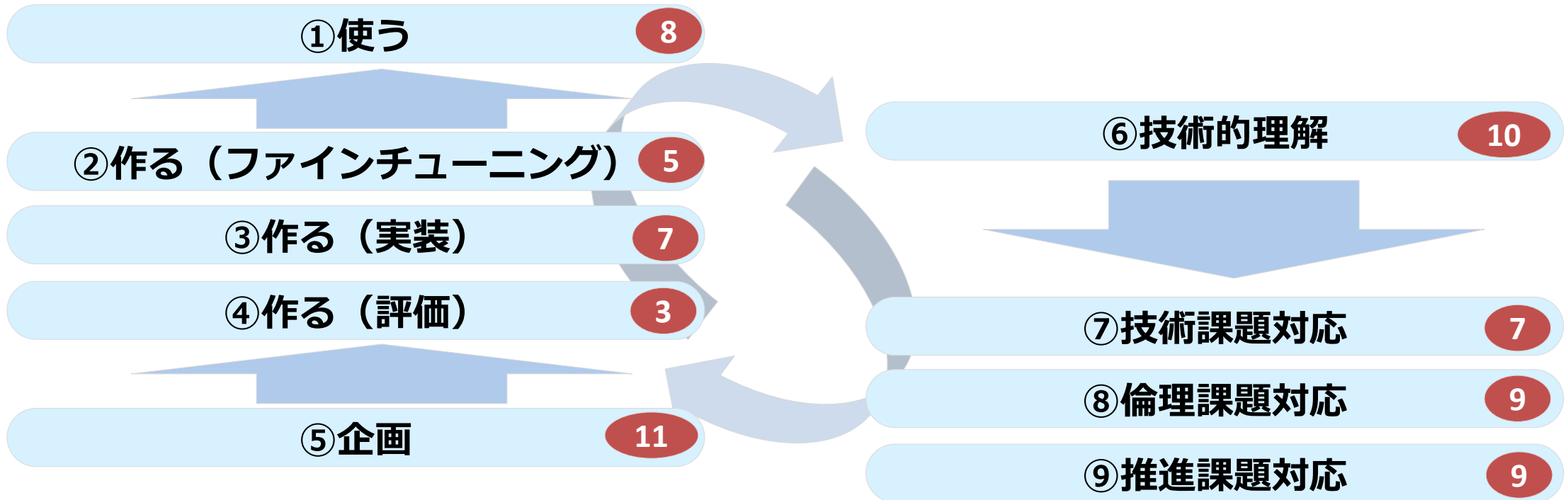
- ✓ 課題解決アイデアの壁打ち
- ✓ 要約、要点抽出
- ✓ ダミーデータの作成
- ✓ オープンデータ収集
- ✓ Pythonコード生成
- ✓ Advanced Data Analysisでデータ分析やグラフ作成
- ✓ PPT資料作成

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

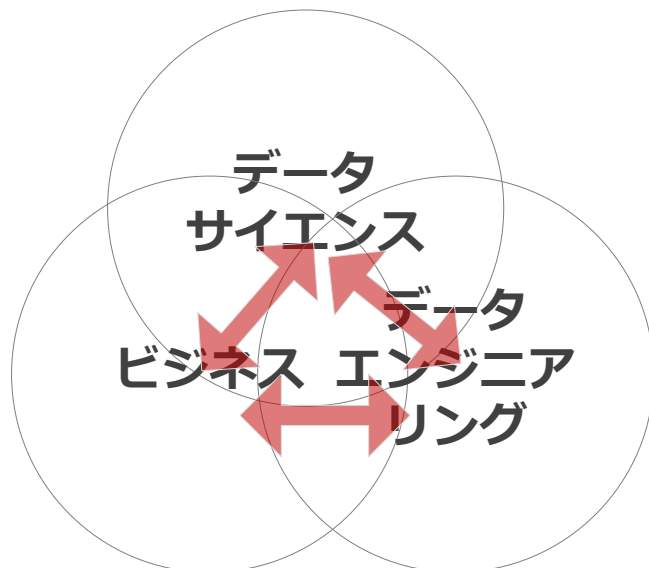
利活用スキル 34

背景理解・対応スキル 35



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

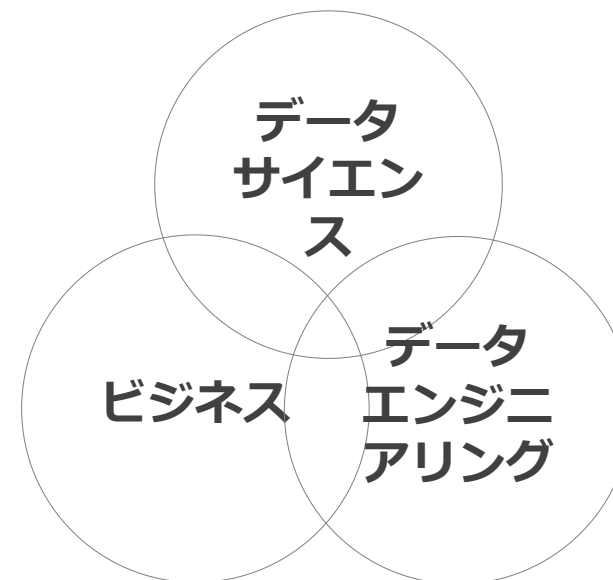


プロジェクト推進役の棟梁クラスは
激しく交差したスキルが必要

★★★
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

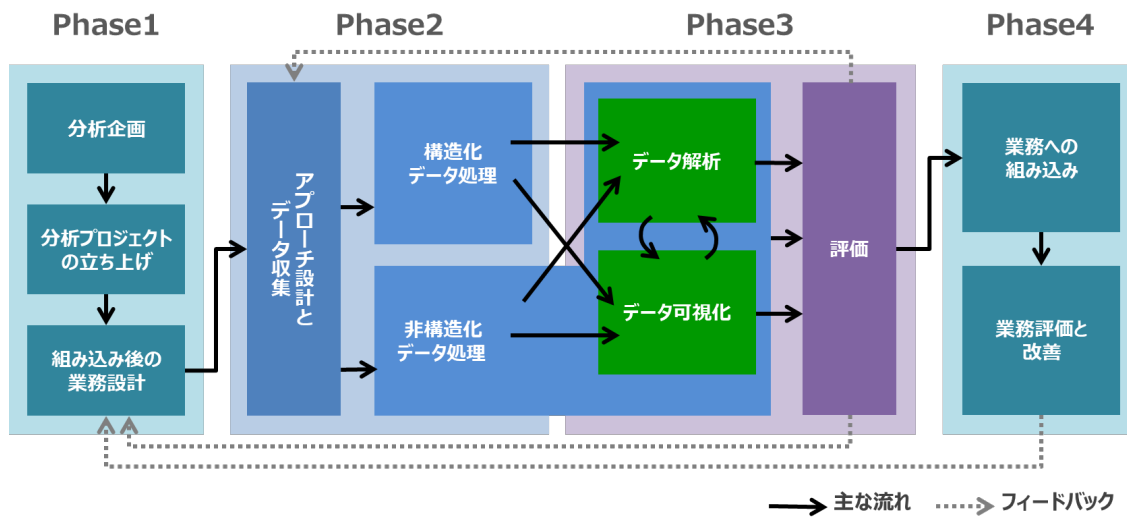
DS：予測モデル構築において頑健性（Robustness）を維持するための具体的な方法を設計、実施できる

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

タスクリストは従来型のDSとAI利活用の2つに

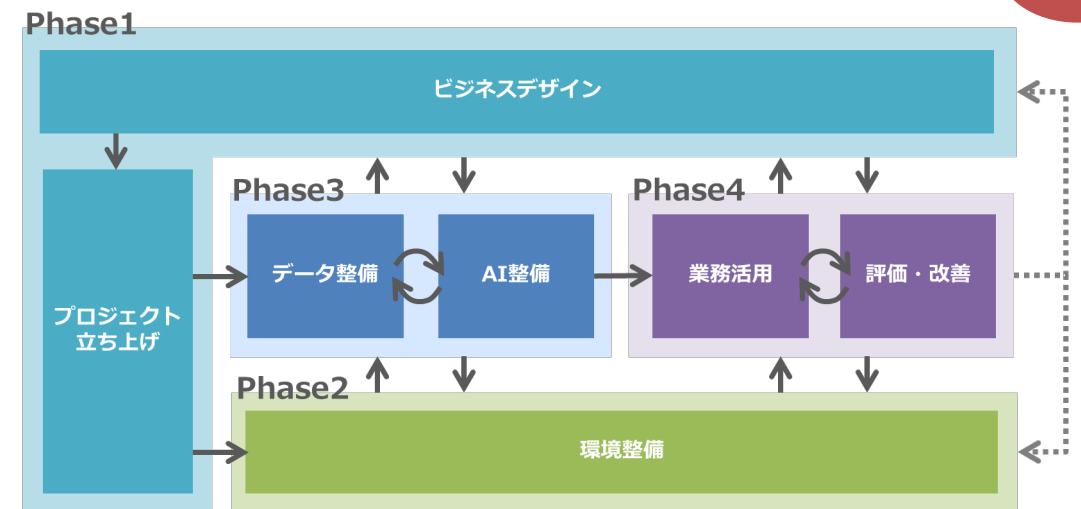
ITSS+ データサイエンス領域のタスクリスト Ver.4

データサイエンス タスクリスト



ITSS+ データサイエンス領域のタスクリストとして
2017年4月に公開

AI利活用タスクリスト



生成AI時代に対応したタスクリストとして
2023年10月に新規定義

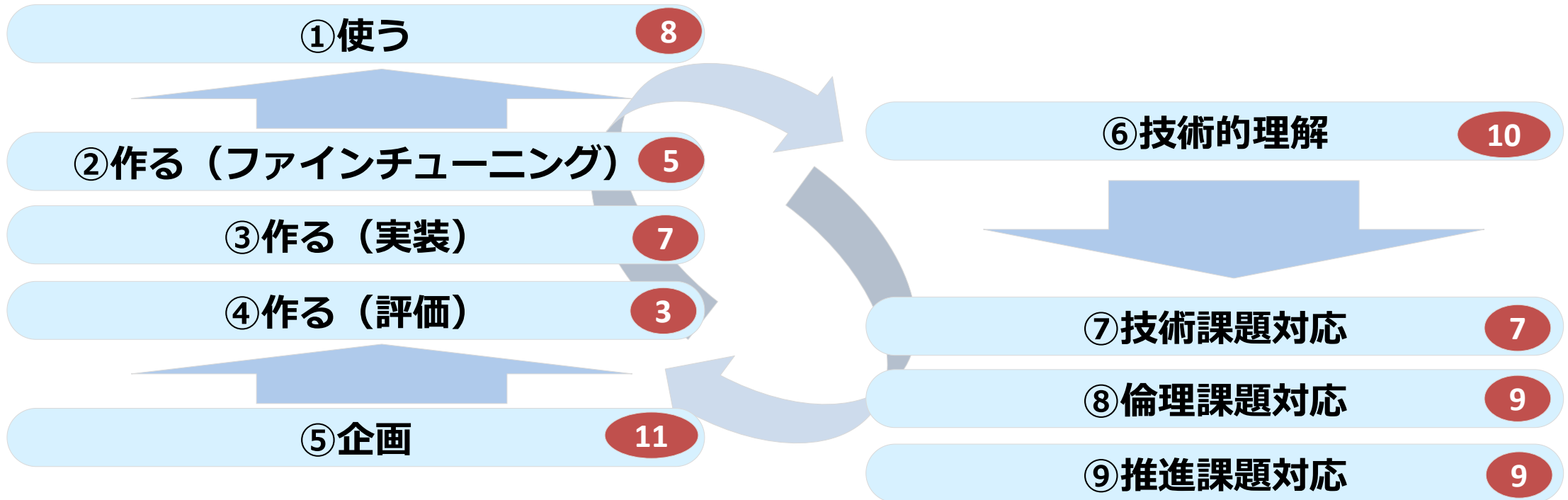
詳細を見てみましょう

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



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

★ 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

★ ★ 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

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



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を理解し、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）、今起き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

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

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

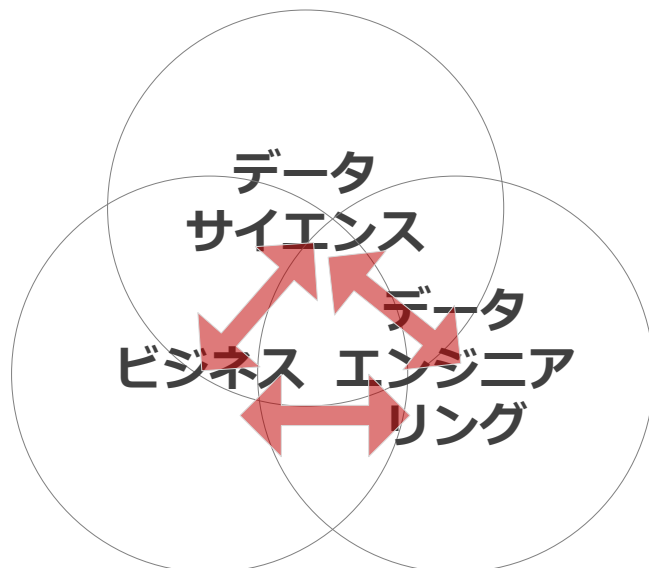
フェーズ	ビジネス	データサイエンス	データエンジニアリング
①使う	3		5
②作る-FT		2	3
③作る-実装	2	1	4
④作る-評価		3	
⑤企画	8	3	
⑥技術的理解	4	6	
⑦技術課題対応	4	3	
⑧倫理課題対応	9		
⑨推進課題対応	8	1	
総計	38	19	12

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

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

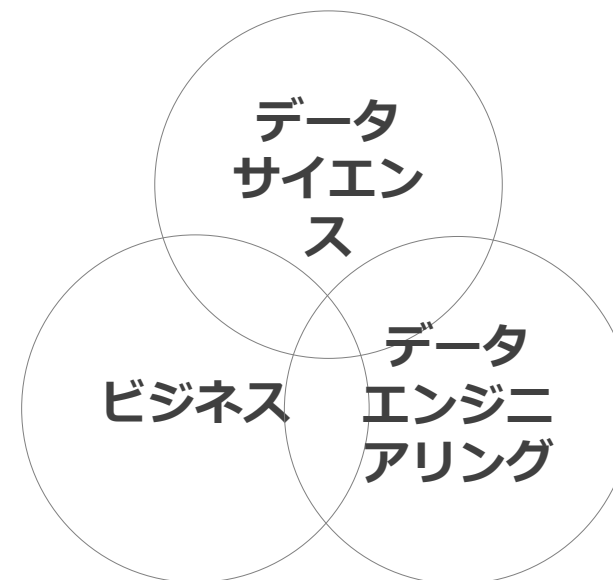


プロジェクト推進役の棟梁クラスは
激しく交差したスキルが必要

★★★
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

アップデート内容のハイライト

全体

- 生成AI活用スキルをそれぞれに追加
- カバー範囲の広いスキルが多く「またがり」スキル増加
- モデル性能を向上させるためのチューニング・リソース管理のスキル等を追加

ビジネス

- AI活用時のビジネス視点を強化
「データ理解」「着想デザイン」スキル追加
「統計情報への正しい理解」をDSより移管
「アプローチ設計」に「生成AI活用」サブカテゴリ追加
- 組織全体にまたがるスキル項目を追加
「組織マネジメント」「契約・権利保護」にスキル追加

データ エンジニア リング

- クラウド環境のさらなる発展
- 利用・構築のスキル
プロンプトエンジニアリング、コーディング支援、ファインチューニング
- 学習済モデル/サービスの開発
利用環境の整備をDSから移動、予測モデル・各種サービスの利用
- 新技術/新スキル
無線通信網を用いたデータ収集、AIを活用したサイバー攻撃への対応

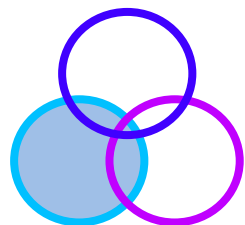
データ サイエンス

- 科学的解析の基礎統合
性質・関係性、アソシエーション分析の名称変更
- モデリング関連スキル→モデル化、モデル利活用
モデル化 / ネットワーク分析（新設）、モデル利活用 / レコメンド
- カテゴリ名称見直し
基礎数学→ 数学的理解、データ加工・自然言語処理の加工スキル→データ準備、シミュレーションデータ同化・最適化→オペレーションズ・リサーチ

タスク リスト

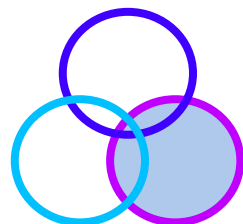
- AI利活用タスクを追加
技術を理解した上でのビジネススキル
ファインチューニングやアライメント
UI/UXの重要性を反映
ガイドラインの策定

スキル表 項目数



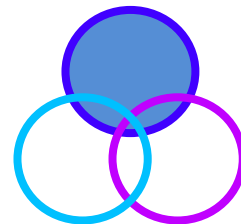
データサイエンス

2015 2017 2019 2021 2023



エンジニアリング

2021 2023



ビジネス

2021 2023

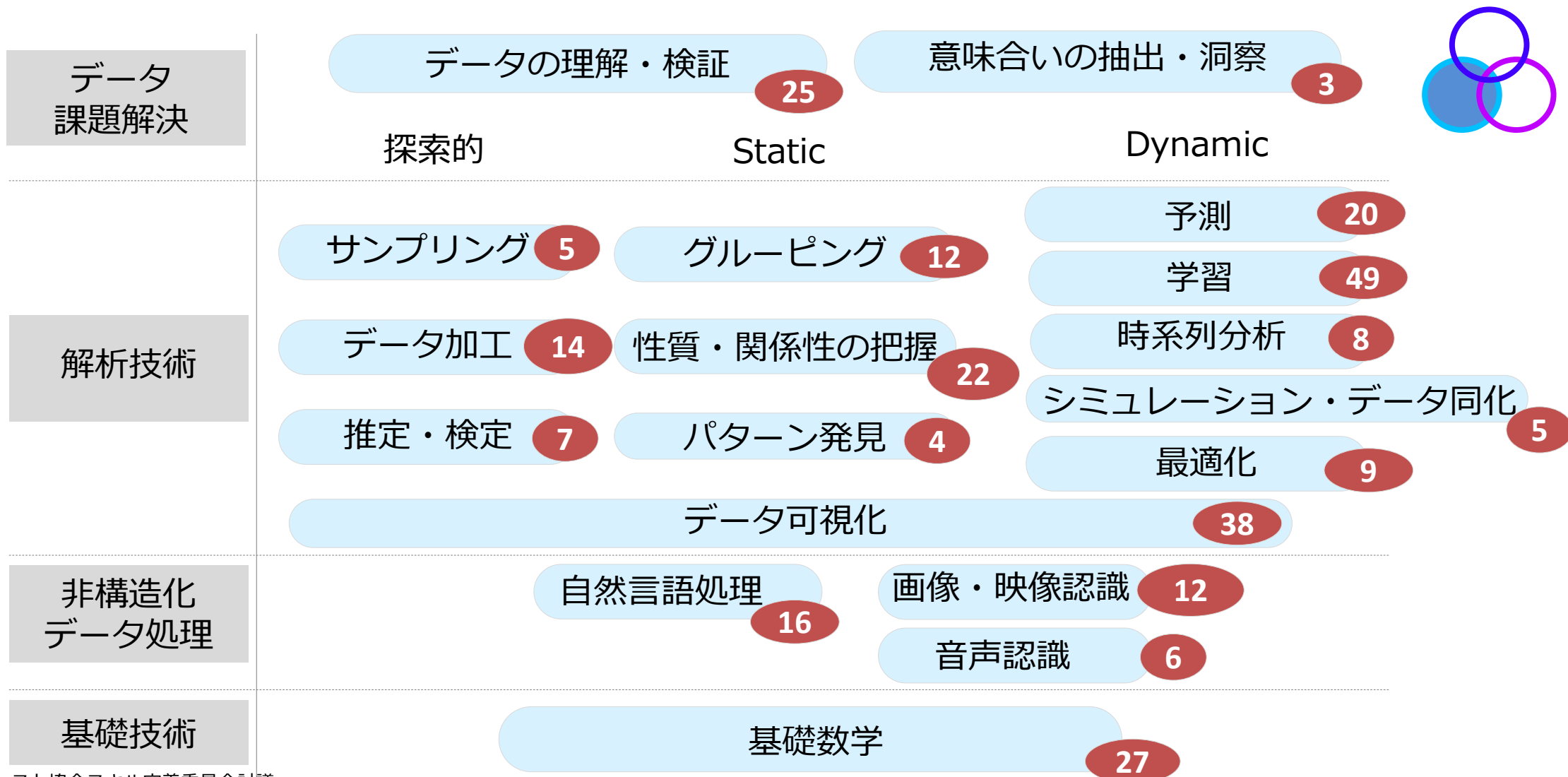
レベル計

2021 2023

★★★ 棟梁レベル (フル)	58	71	64	62	72	37	39	42	37	45	44	38	42	47	60	139	148	149	146	177
★★ 一人前レベル (アソシエート)	68	90	121	120	133	43	52	63	65	72	49	42	49	56	65	160	184	232	241	270
★ 見習いレベル (アシスタント)	54	67	86	100	104	39	38	39	57	65	30	20	22	28	34	123	125	149	185	203
領域計	180	228	271	282	309	119	129	144	159	182	123	100	113	131	159	422	457	528	572	650

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

Ver.4



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

Ver.5

AI活用

生成

14

データ
課題解決

データの理解・検証

21

- ✓ 生成スキルを中心としたAI活用スキルを追加
- ✓ 既存のカテゴリについては整理・統合を実施

探索的

Static

Dynamic

解析技術

データ準備

19

科学的解析の基礎

51

モデル化

106

モデル利活用

5

パフォーマンスリサーチ

14

データ可視化

38

非構造化
データ技術

非構造化データ処理

32

基礎技術

数学的理解

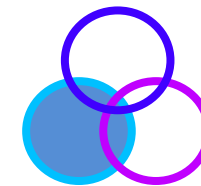
9

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

アップデートの詳細と例 ①データサイエンス

カテゴリを目的軸、サブカテゴリを技術軸として再編成

- ① 生成AIに関するスキル群を追加し、これらを束ねる「生成」カテゴリを追加
- ② 解析の科学的実践に必要な基本的スキルを「科学的解析の基礎」として統合
- ③ モデリング関連スキルを「モデル化」、「モデル利活用」として統合
- ④ その他、統合や名称変更、領域移動による変更
- ⑤ スキルカテゴリの順番をデータ分析の手順に沿うよう変更



参考) これまでのデータサイエンスにおけるカテゴリ・サブカテゴリ構成（2021年Ver.4 時点）

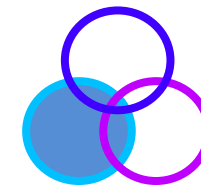
スキルカテゴリ	サブカテゴリ
基礎数学	統計数理基礎
	線形代数基礎
	微分・積分基礎
	集合論基礎
データの理解・検証	統計情報への正しい理解
	データ確認
	俯瞰・メタ思考
	データ理解
	データ粒度
意味合いの抽出、洞察	洞察
予測	回帰・分類
	評価
推定・検定	推定・検定

スキルカテゴリ	サブカテゴリ
グルーピング	グルーピング
	異常検知
性質・関係性の把握	性質・関係性の把握
	グラフィカルモデル
	因果推論
サンプリング	サンプリング
データ加工	データクレンジング
	データ加工
	特徴量エンジニアリング
データ可視化	方向性定義
	軸出し
	データ加工
	表現・実装技法
	意味抽出

スキルカテゴリ	サブカテゴリ
時系列分析	時系列分析
学習	機械学習
	深層学習
	強化学習
自然言語処理	自然言語処理
画像・映像認識	画像認識
	映像認識
音声認識	音声認識
パターン発見	パターン発見
シミュレーション・データ同化	シミュレーション・データ同化
最適化	最適化

アップデートの詳細と例 ①データサイエンス

① 生成AIに関するスキル群を追加し、これらを束ねる「生成」カテゴリを追加



Ver.5 カテゴリ構成

スキルカテゴリ	サブカテゴリ
生成	大規模言語モデル
	画像生成モデル
	オーディオ生成モデル

生成カテゴリの中では、大規模言語モデル関連のスキルが特に多いが、これは以下の背景を踏まえたものとなっている

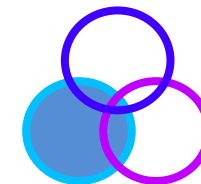
- 言語モデルはその他非構造化データモデルに対してカバーしているタスク領域が広い
- 様々な学習済みモデルやそれに対するFT手法などの登場により必要とされるスキルの幅が広がっている
- 画像やオーディオに対して、言語は様々なビジネスシーンで活用されやすい
- 画像やオーディオはそれ自身が創造物であるのに対し、言語モデルは出力結果が真実であることを期待して使われるため、Hallucinationに関する理解が重要となる

生成AIを使いこなすためのスキル群を新たに追加

- ✓ Hallucinationに関する理解
- ✓ 大規模言語モデル（LLM）を使うことの（独自モデルに対する）メリット
- ✓ 大規模言語モデルに対する評価
- ✓ 大規模言語モデル（LLM）とそれ以前の言語モデルの違い理解や導入推進
- ✓ オープンソースLLMに対するFine Tuning
- ✓ 基盤モデルとその他モデルとの違いに対する理解
- ✓ 拡散モデルに対する理解
- ✓ 画像生成モデルにおける評価の考え方
- ✓ オーディオ生成モデルにおけるエンコーディング技術の理解
- • • など

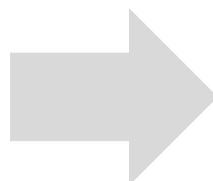
アップデートの詳細と例 ①データサイエンス

② 解析の科学的実践に必要な基本的スキルを「科学的解析の基礎」として統合



Ver.4 カテゴリ構成

スキルカテゴリ	サブカテゴリ
基礎数学	統計数理基礎
	線形代数基礎
	微分・積分基礎
	集合論基礎
意味合いの抽出、洞察	洞察
推定・検定	推定・検定
性質・関係性の把握	性質・関係性の把握
	グラフィカルモデル
	因果推論
パターン発見	パターン発見

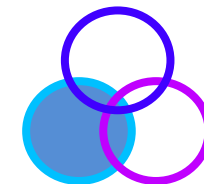


Ver.5 カテゴリ構成

スキルカテゴリ	サブカテゴリ
科学的解析の基礎	統計数理基礎
	洞察
	性質・関係性 ※旧性質・関係性の把握
	推定・検定
	アソシエーション分析 ※旧パターン発見
	因果推論

- ✓ データ分析を科学的に行うにあたり基礎となるスキル群を統合
- ✓ 特に、モデル化を行う前段となる解析スキルが中心

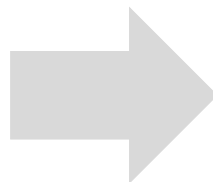
アップデートの詳細と例 ①データサイエンス



③ モデリング関連スキルを「モデル化」、「モデル利活用」として統合

Ver.4 カテゴリ構成

スキルカテゴリ	サブカテゴリ
予測	回帰・分類
	評価
グルーピング	グルーピング
	異常検知
性質・関係性の把握	性質・関係性の把握
	グラフィカルモデル
	因果推論
時系列分析	時系列分析
学習	機械学習
	深層学習
	強化学習



Ver.5 カテゴリ構成

スキルカテゴリ	サブカテゴリ
モデル化	回帰・分類
	統計的評価（旧評価）
	機械学習
	深層学習
	強化学習
	時系列分析
	クラスタリング（旧グルーピング）
	グラフィカルモデル
	ネットワーク分析（新設）
モデル利活用	異常検知
	レコメンド（新設）

- ✓ 「モデル化」にまつわるカテゴリと、応用的な「モデル利活用」に統合
- ✓ 機械学習にはデータの質に関する Data Centric AI関連スキルを追加
- ✓ グラフデータに基づくネットワーク分析やレコメンドをサブカテゴリとして新設

アップデートの詳細と例 ①データサイエンス

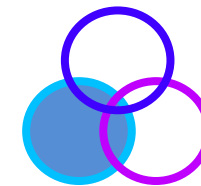
④ その他、統合や名称変更、領域移動による変更

Ver.4 カテゴリ構成

スキルカテゴリ	サブカテゴリ
基礎数学	統計数理基礎
	線形代数基礎
	微分・積分基礎
	集合論基礎
データの理解・検証	統計情報への正しい理解
	データ確認
	俯瞰・メタ思考
	データ理解
	データ粒度
サンプリング	サンプリング
データ加工	データクレンジング
	データ加工
自然言語処理	特徴量エンジニアリング
	自然言語処理
画像・映像認識	画像認識
	映像認識
音声認識	音声認識
シミュレーション・データ同化	シミュレーション・データ同化
最適化	最適化

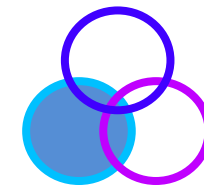
Ver.5 カテゴリ構成

スキルカテゴリ	サブカテゴリ
数学的理解	統計数理基礎 → 「科学的解析の基礎」に移動
	線形代数基礎
	微分・積分基礎
	集合論基礎
データの理解・検証	統計情報への正しい理解 → ビジネスカに移動
	データ確認
	俯瞰・メタ思考
	データ理解
	データ粒度
データ準備	サンプリング
	データクレンジング
	データ加工
	特徴量エンジニアリング
非構造化データ処理	自然言語処理
	画像認識
	映像認識
	音声認識
オペレーションズリサーチ	シミュレーション・データ同化
	最適化



アップデートの詳細と例 ①データサイエンス

⑤ スキルカテゴリの順番をデータ分析の手順に沿うように変更



ver.5 カテゴリ構成

スキルカテゴリ
数学的理解
科学的解析の基礎
データの理解・検証
データ準備
データ可視化
モデル化
モデル利活用
非構造化データ処理
生成
オペレーションズリサーチ

分析
の
流
れ

データ分析前の基礎理解

- ✓ 数学的理解
- ✓ 科学的解析の基礎

モデル化前のデータ理解・処理

- ✓ データの理解・検証
- ✓ データ準備
- ✓ データ可視化

モデルを活用したデータ分析

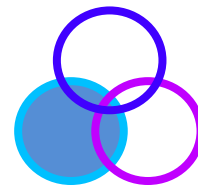
- ✓ モデル化
- ✓ モデル利活用
- ✓ 非構造化データ処理
- ✓ 生成
- ✓ オペレーションズリサーチ

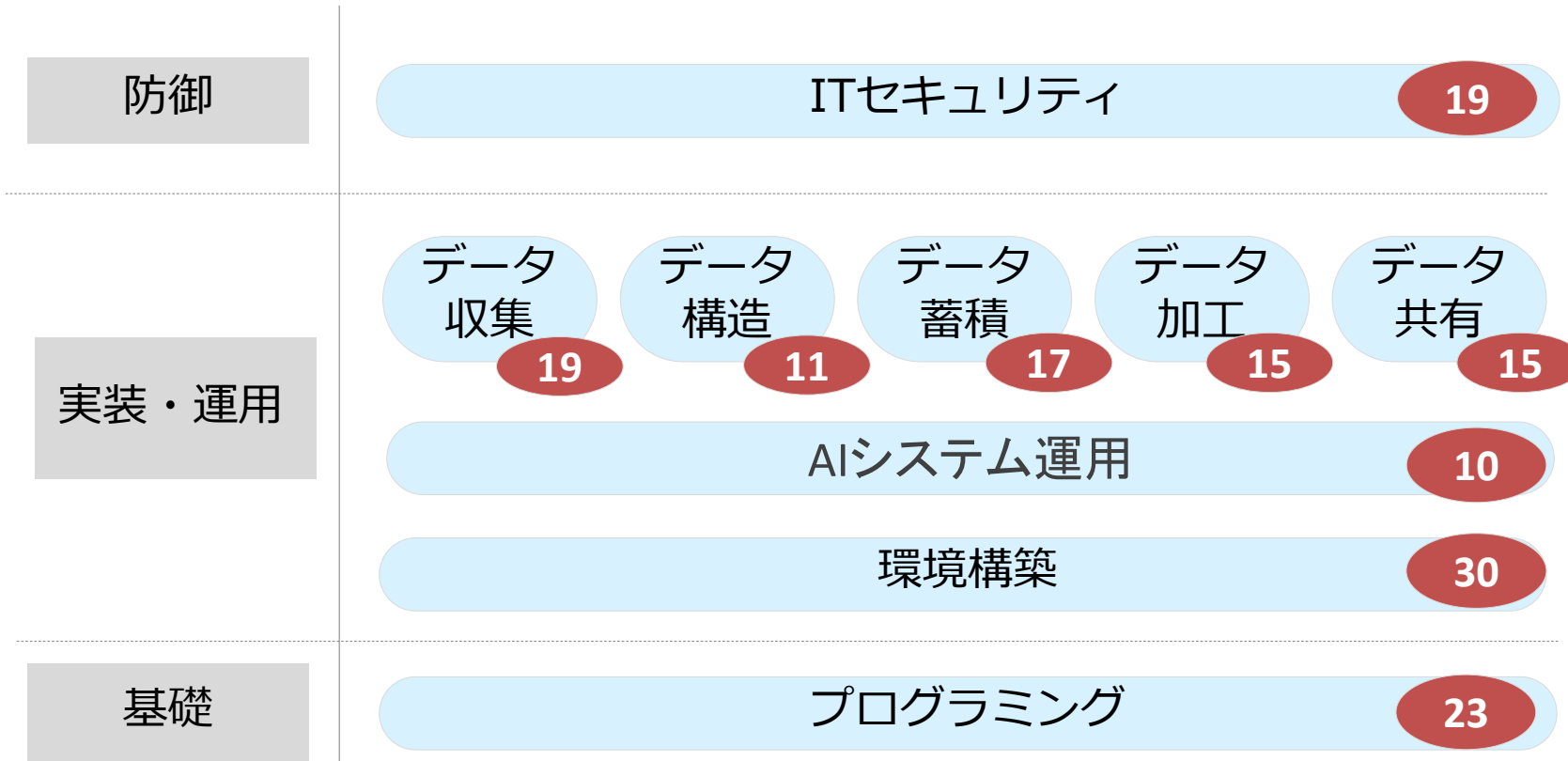
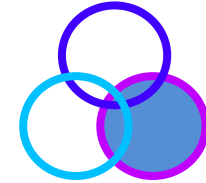
アップデートの詳細と例 ①データサイエンス

再編成後のカテゴリ構成全体像（ver.5）

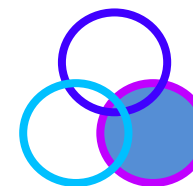
スキルカテゴリ	サブカテゴリ
数学的理解	線形代数基礎
	微分・積分基礎
	集合論基礎
科学的解析の基礎	統計数理基礎
	洞察
	性質・関係性
	推定・検定
	アソシエーション分析
	因果推論
データの理解・検証	データ確認
	俯瞰・メタ思考
	データ理解
	データ粒度
データ準備	サンプリング
	データクレンジング
	データ加工
	特徴量エンジニアリング
データ可視化	方向性定義
	軸出し
	データ加工
	表現・実装技法
	意味抽出

スキルカテゴリ	サブカテゴリ
モデル化	回帰・分類
	統計的評価
	機械学習
	深層学習
	強化学習
	時系列分析
	クラスタリング
	グラフィカルモデル
	ネットワーク分析
モデル利活用	異常検知
	レコメンド
非構造化データ処理	自然言語処理
	画像認識
	映像認識
生成	音声認識
	大規模言語モデル
	画像生成モデル
オペレーションズリサーチ	オーディオ生成モデル
	シミュレーション・データ同化
	最適化





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



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

アップデートの詳細と例 ②エンジニアリング

① クラウド環境のさらなる発展

クラウド環境で提供される新機能の利活用スキル

システム設計

運用管理環境

アーキテクチャ設計

コンテナオーケストレーション
/マイグレーション

分析プログラム

クラウド統合開発環境

MLOps

特徴量管理

② 新カテゴリ「生成AI」追加

生成AI環境の利用や構築に必要なスキル

プロンプト
エンジニアリング

コーディング
支援

生成AIの
技術利用

ファイン
チューニング

生成AI開発

③ 学習済モデル/サービスの利用

学習済モデル/サービス利用環境整備（DSからの移動）

DSからの移動

画層、音声、自然言語など

分析サービス利用

予測モデルや各種サービス

ソースレビュー機能

レビューや性能改善への利用

④ 新技術/新スキル

通信技術

自動運転や遠隔操作などで用いられる無線通信網をもちいたデータ収集

攻撃と防御手法

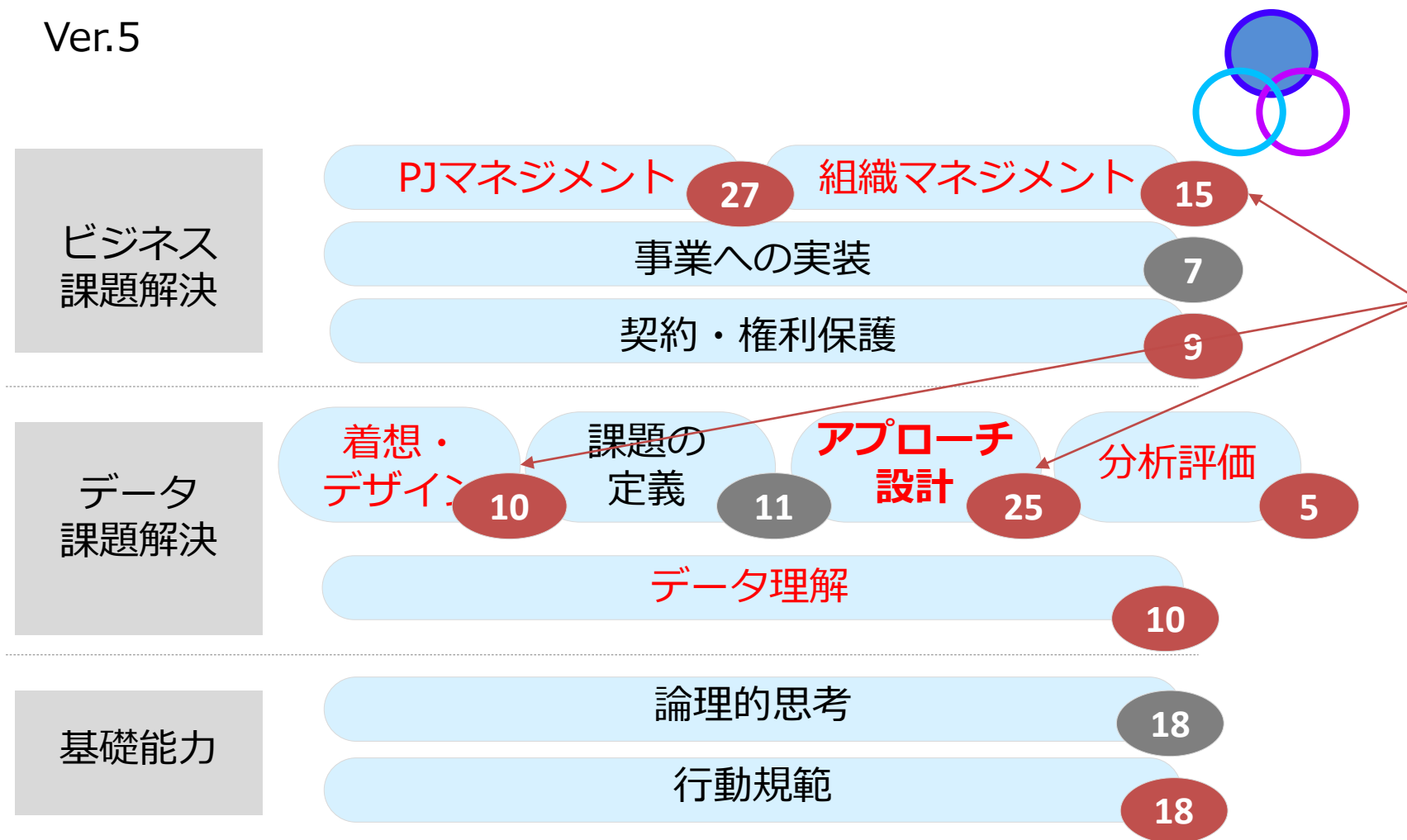
AIを活用したサイバー攻撃防御ソリューションの評価・導入

Ver.4



資料：データサイエンティスト協会スキル委員会討議

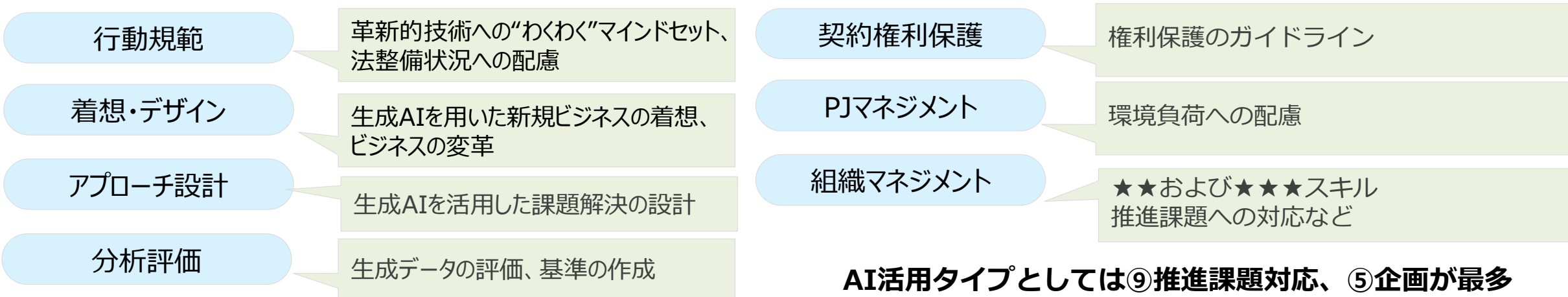
Ver.5



ビジネス・技術的な背景理解は
もちろん、そこから生じる各課題
(技術・倫理・組織的な推進等)
について、経営・組織・仕組みに
提言するといった、組織全体にま
たがる取り組みに対して、スキル
項目を追加

アップデートの詳細と例 ③ ビジネス

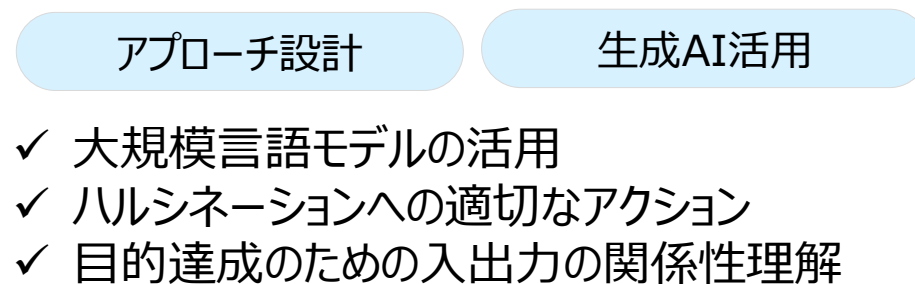
① 生成AI活用スキル多数の新規追加 新規追加スキル23、全てAI利活用スキル



② DSスキルからの移管、またがりスキルの増加

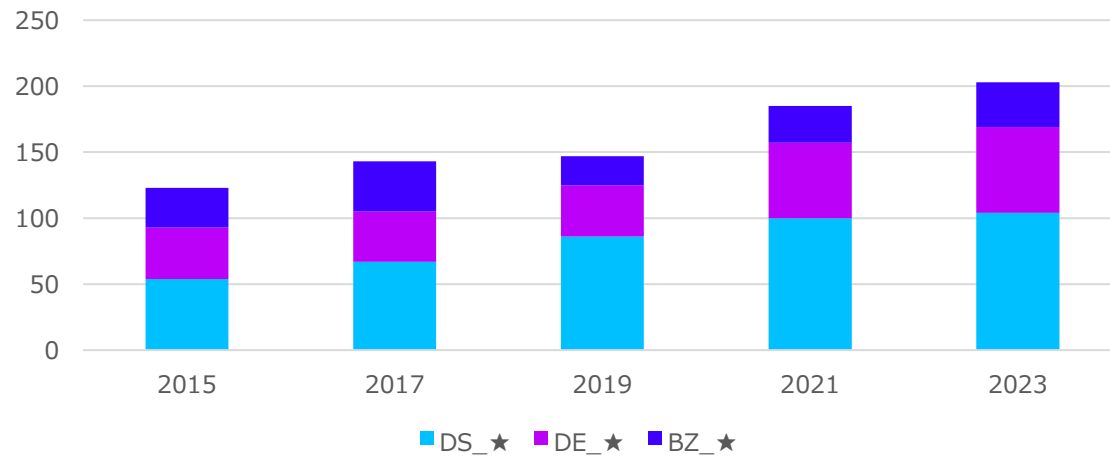
- ✓ データ理解：統計情報の正しい理解（4スキル）
- ✓ DSまたがりスキル 14→24
- ✓ DSまたがりスキル 13→19
- ✓ DS/DEまたがりスキル 30→41

③ 「アプローチ設計」「生成AI活用」サブカテゴリ追加

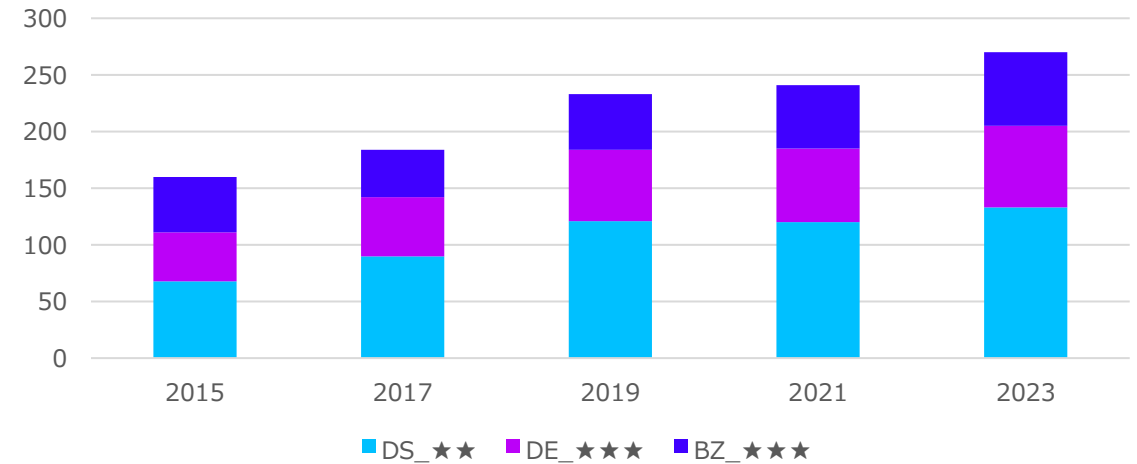


Ver.1～Ver.5振り返り：スキルレベル

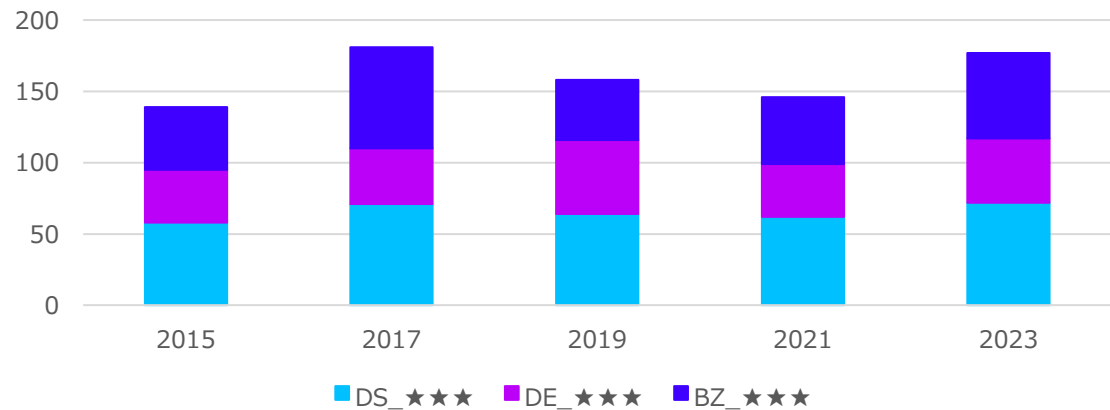
★ 見習いレベル スキル数推移



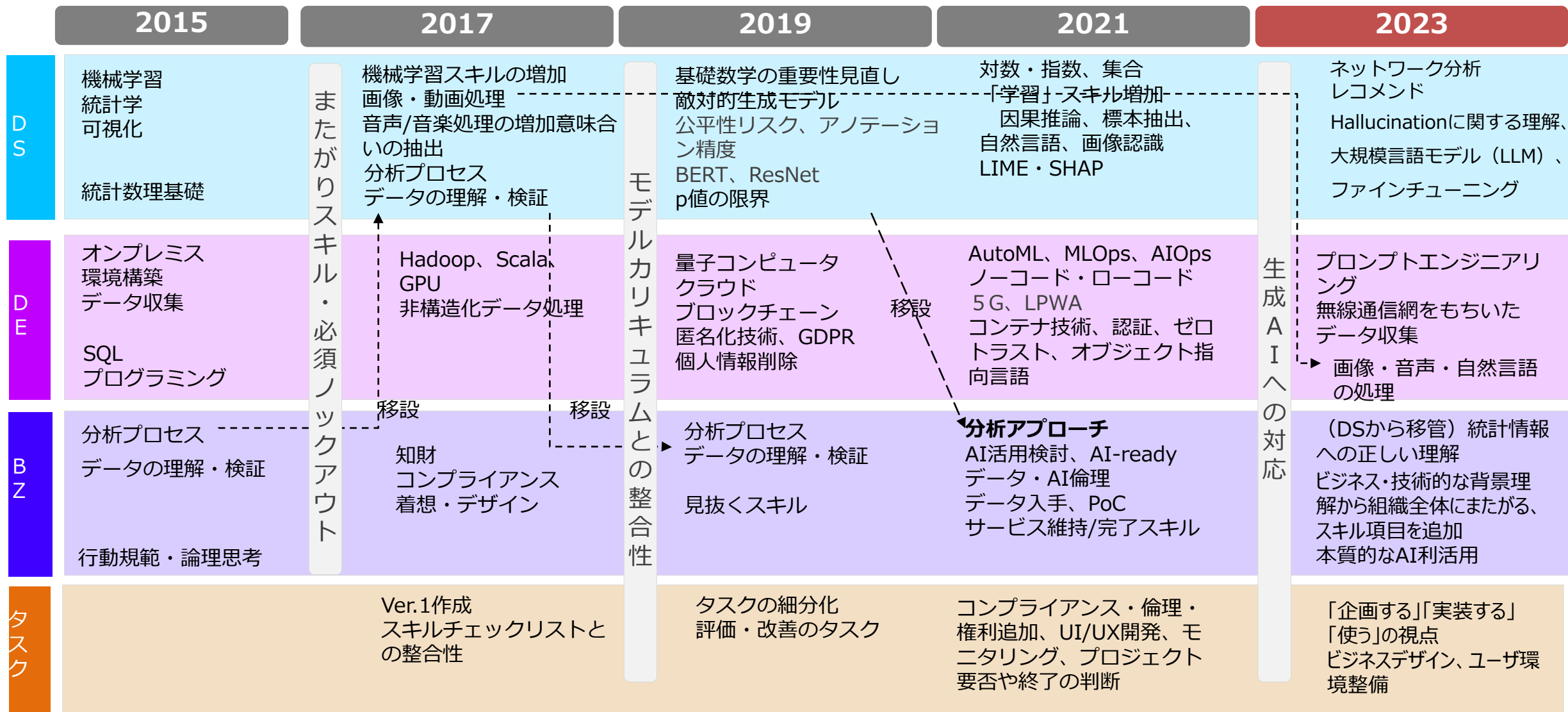
★★ 一人前レベル スキル数推移



★★★ 棟梁レベル スキル数推移



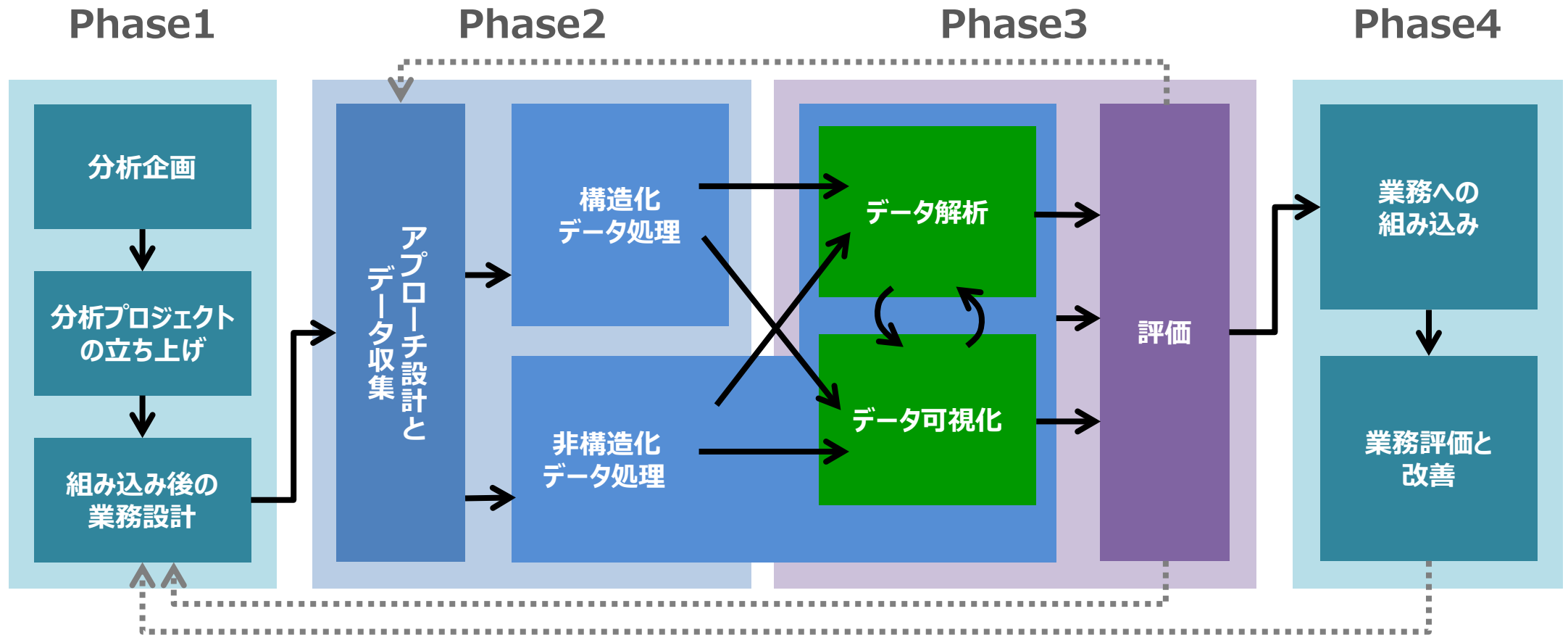
Ver.1～Ver.5振り返り：主要な構造



2021年度版タスクリスト (ver.3)

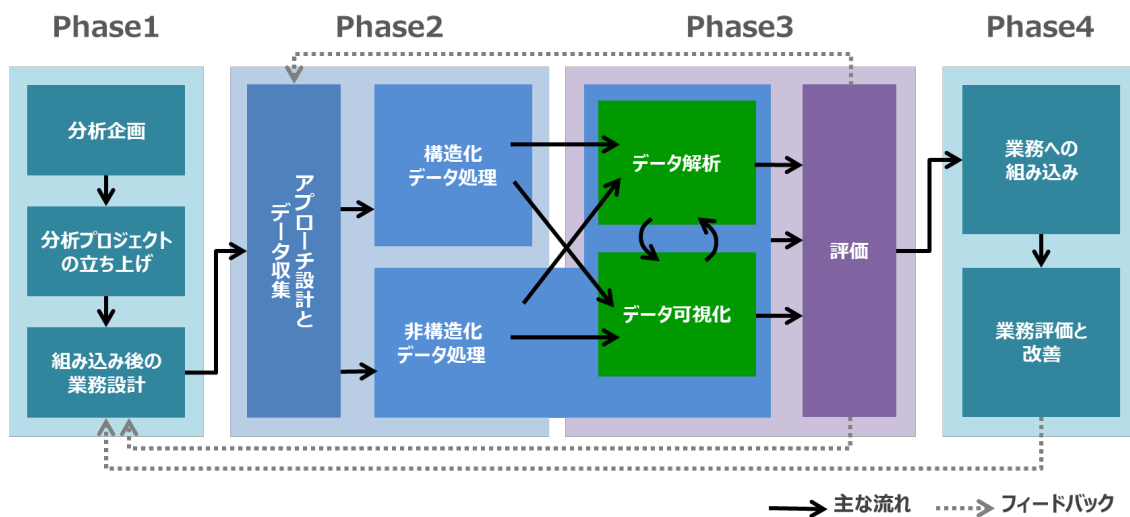
ITSS+ データサイエンス領域のタスクリスト

→ 主な流れ
.....> フィードバック



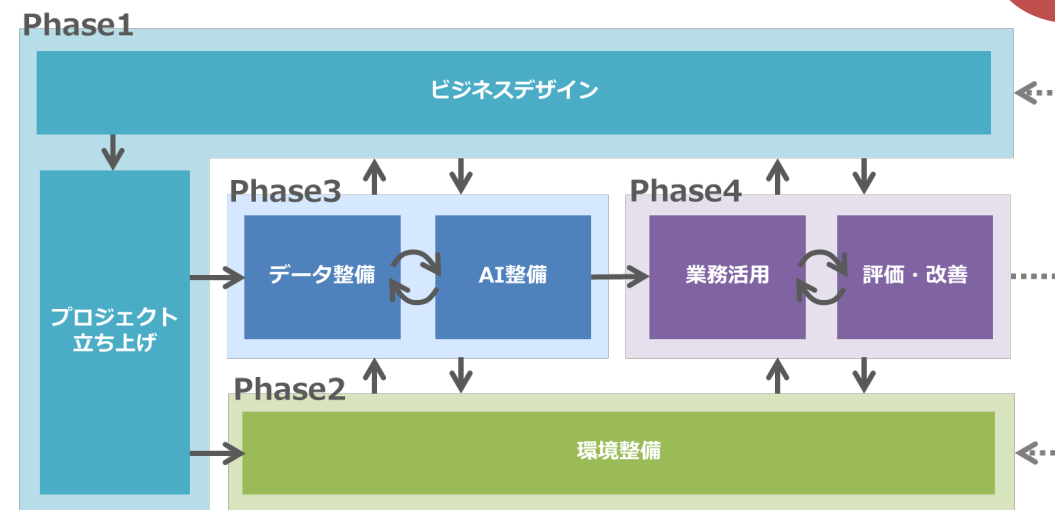
ITSS+ データサイエンス領域のタスクリスト Ver.4

データサイエンス タスクリスト



ITSS+ データサイエンス領域のタスクリストとして
2017年4月に公開

AI活用タスクリスト



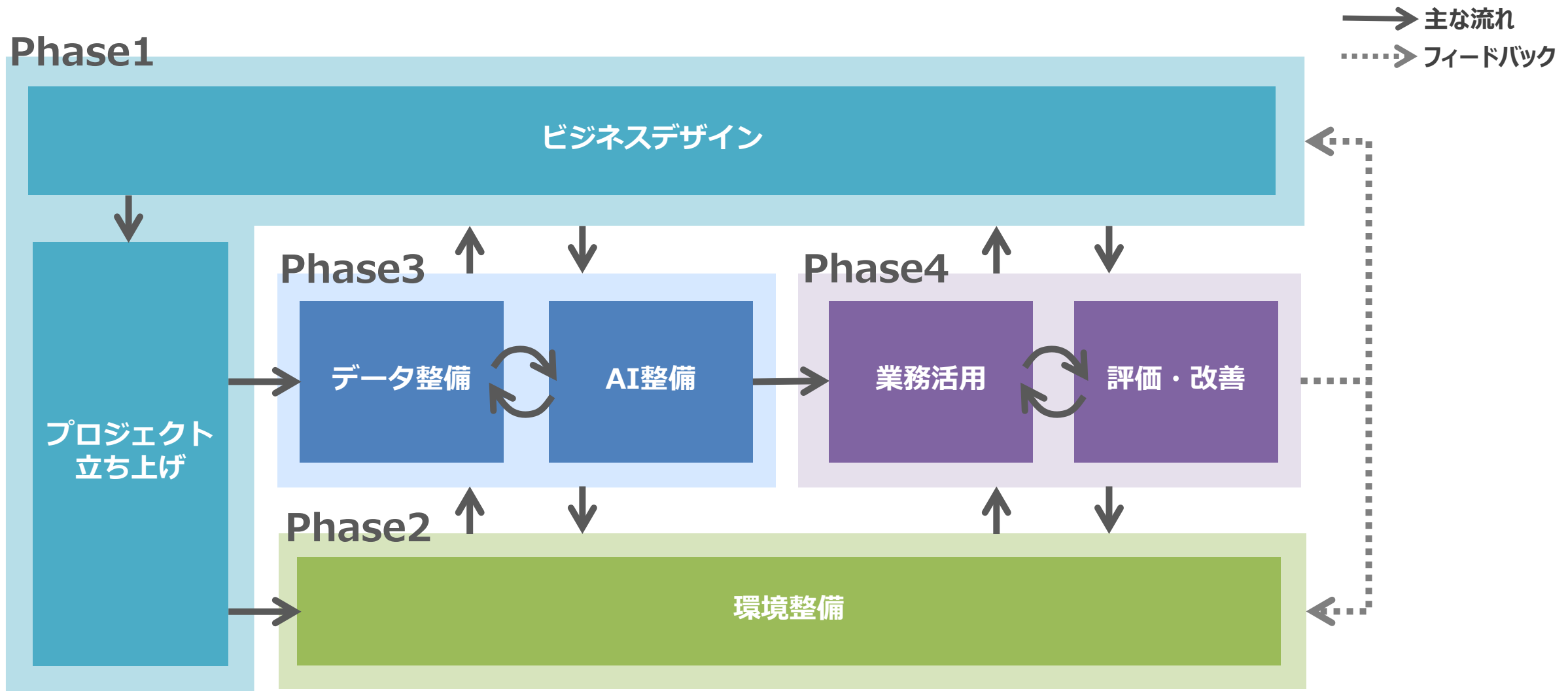
生成AI時代に対応したタスクリストとして
2023年10月に新規定義

AI活用タスクリスト作成の背景と理由

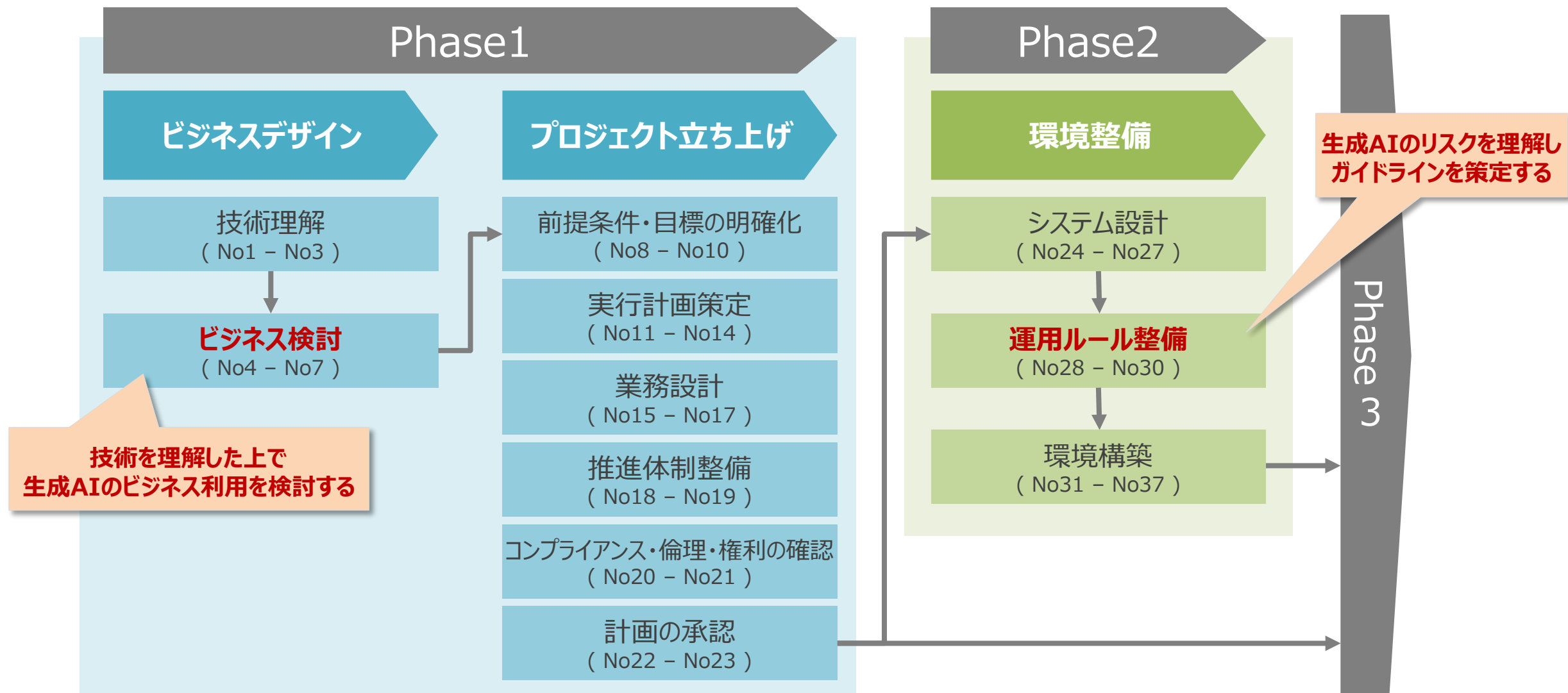
- 生成AI(大規模言語モデル・拡散モデル)の登場によってデータサイエンティストの業務も大きく変化すると考えられる
- 生成AIの利用におけるタスクを「**企画する**」という視点や「**実装する**」という視点、「**使う**」という視点で整理
- 生成AIの利用においては「**ビジネスデザイン**」や「**ユーザ環境整備**」が重要となる
- 既存のデータサイエンティストタスクリストとは異なるタスクも多くなるため独立した**AI活用タスクリスト**として作成した



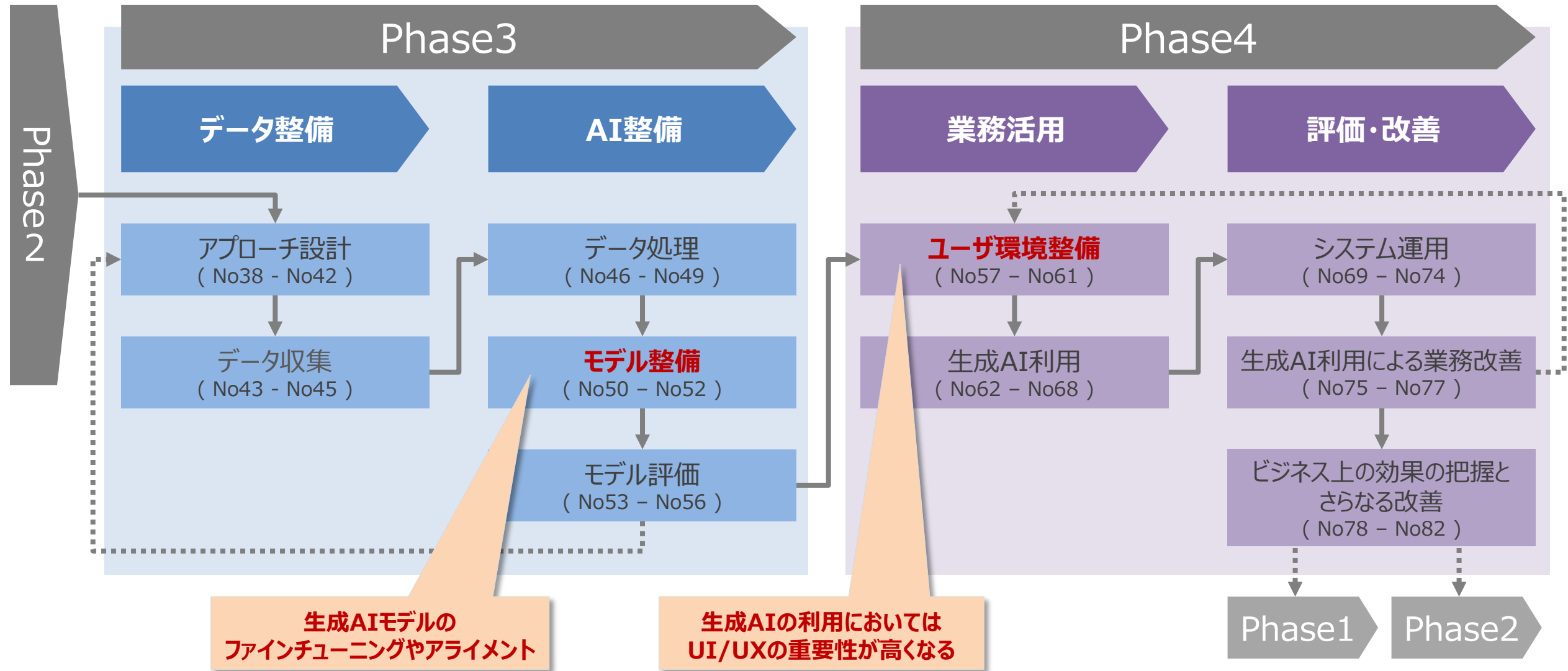
AI活用タスクリスト<構造図>



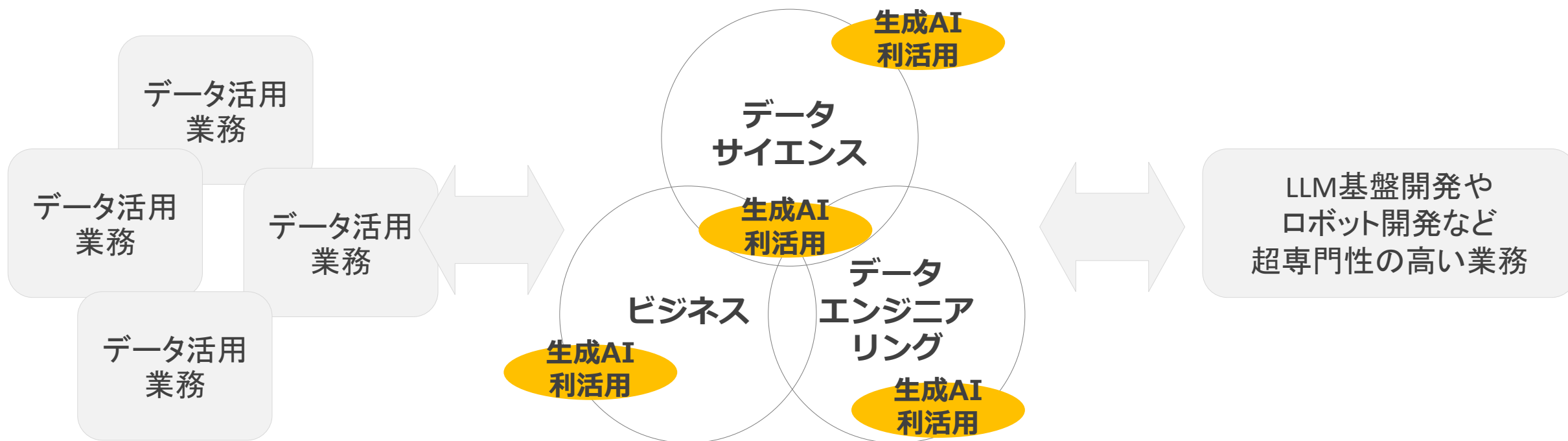
AI活用タスクリスト<Phase1～2>



AI利活用タスクリスト<Phase3~4>



生成AI時代のデータサイエンティスト、データプロフェッショナルに必要な人材像や仕事はどう変わっていくのでしょうか？



今後の公開予定・更新予定

シンポジウム直後

本日の講演資料を協会HPからダウンロード可能に

11月1日
(予定)

データサイエンティスト協会HP、IPAサイトで公開

- ・スキルチェックリスト ver.5
- ・タスクリスト ver.4

2024年1月目途

概説書のアップデート IPAサイト、協会HPからお知らせ

2024年6月目途

DS検定バージョンアップ。生成AI利活用スキルも追加

