

教育環境分科会 2023年度会合

「生成AIを活用した教育実践の現状と課題」

文系大学のデータサイエンス教育と生成AI ～これからのデータリテラシー教育～

高崎商科大学 商学部

講師 中鉢直宏

自己紹介

- 名前：中鉢直宏（ちゅうばちなおひろ）
- 所属 商学部 講師
 - 2021年～ 高崎商科大学
 - 2016年 帝京大学 高等教育開発センター
 - 2014年 島根大学 教学企画IR室
 - 2009年 青山学院大学 情報メディアセンター
- その他
 - (株) アイセス（現：株式会社アイ・ラーニング）ITインストラクター
- 連絡
 - n-chubachi@uv.tuc.ac.jp
- 研究分野
 - 情報教育、情報システム、教育工学

データに関する教育の背景

- 平成29年度【情報通信白書】（総務省）「データ主導経済と社会変革」
- データ主導経済（data-driven economy）
 - データは「21世紀の石油」
 - 多種多様なデータの生成・収集・流通・分析・活用を徹底的に図る
 - 社会経済活動を再設計し、社会の抱える課題の解決
 - データの利活用
 - △データを多く集めること
 - ◎データから取り出される様々な意味や知見を活用

大学に求められるデータサイエンス教育

- 「**大学の数理・データサイエンス教育強化方策について**」
 - Society 5.0の実現
 - 理系学生を対象としたデータサイエンスの専門知識習得を目的とした教育強化策
- 「**理数系人材育成に向けた取組について**」（経済産業省）
 - 「我が国の産業活動を活性化させるために必要な数理・データサイエンスの基礎的素養を持ち、課題解決や価値創出につなげられる人材育成が必要不可欠」
- 大学へのデータ教育の施策
 - 平成28年に「大学の数理及びデータサイエンスに係る教育強化」の拠点校が選定
 - 大学において専門分野の枠を超えた全学的な数理・データサイエンス教育機能を持つセンターの整備
 - 実践的な教育内容・方法の採用や標準カリキュラム・教材の作成が実施
 - 全国の大学への展開・普及

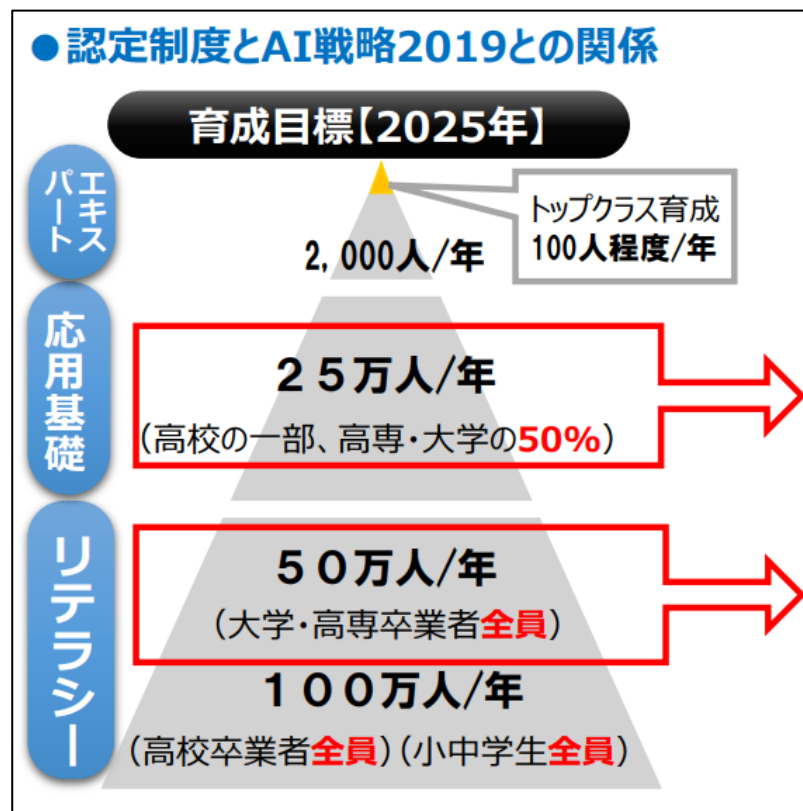
「AI戦略2019」 および 「AI戦略2022」

• 2025年までの 人材育成目標を設定

- 「文理を問わず、全ての大学・高専生（約50万人卒/年）が、**課程にて初級レベルの数理・データサイエンス・AI**を習得」
- 「大学生、社会人に対するリベラルアーツ教育の充実（一面的なデータ解析の結果やAIを鵜呑みにしないための批判的思考力の養成も含む）」
- 「AIに関連する産業競争力強化や技術開発等についての総合戦略」
 - 文系を含む大学教育全体における教養としてのデータサイエンス・AI教育の拡大

教育機関におけるデータ教育の認定制度

- 「数理・データサイエンス・AI教育プログラム認定制度（リテラシーレベル）」
（内閣府、経済産業省、および文部科学省）
 - リテラシーレベルの5つの審査項目
 - 大学改革推進等補助金（デジタル活用高度専門人材育成事業）の要件



- ①リテラシー：約50万人/年（全ての大学・高専生）
- ②応用基礎：約25万人/年
- ③エキスパート：約2,000人/年
- ④トップ：100人程度

https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00002.htm

高崎商科大学をはじめ多くの文系大学が取得 6

データ主導社会の人材育成のイメージ

情報教育

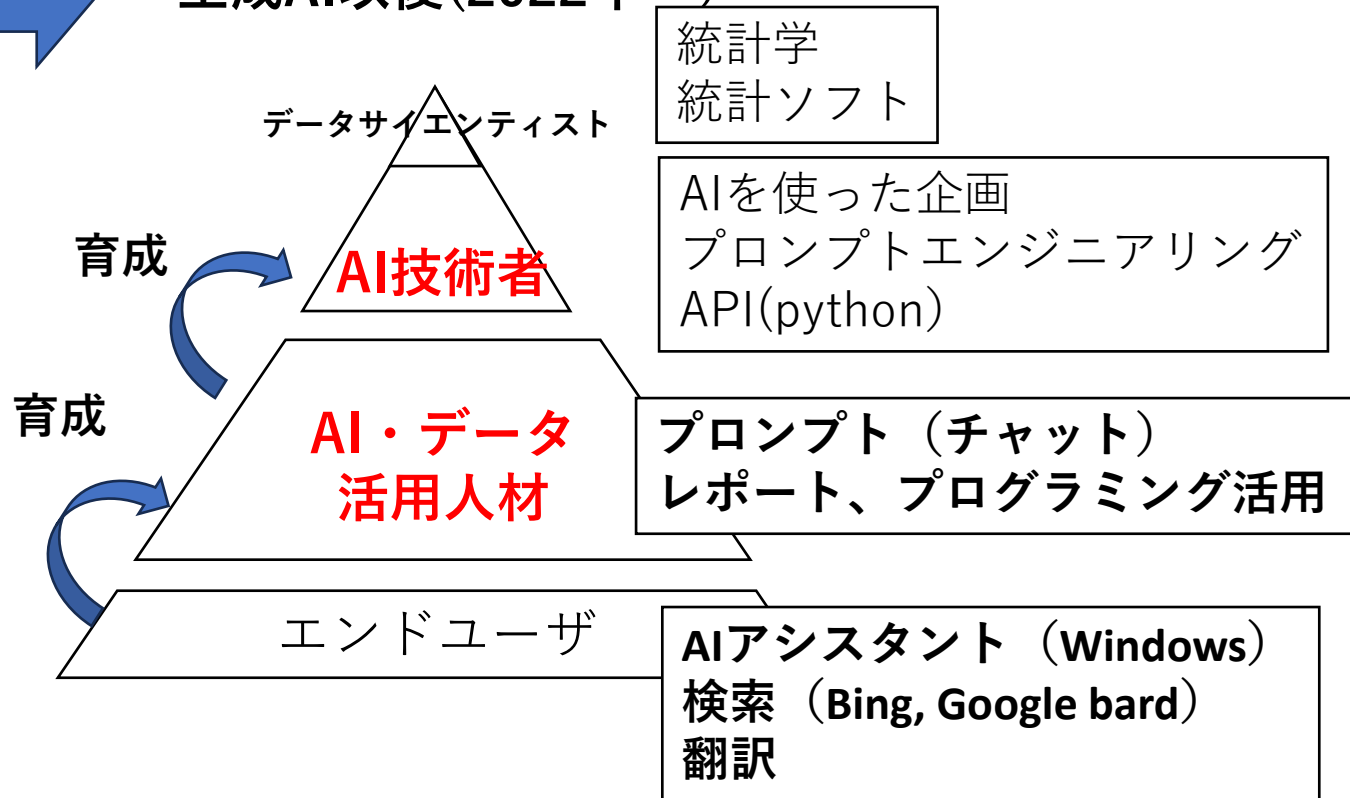
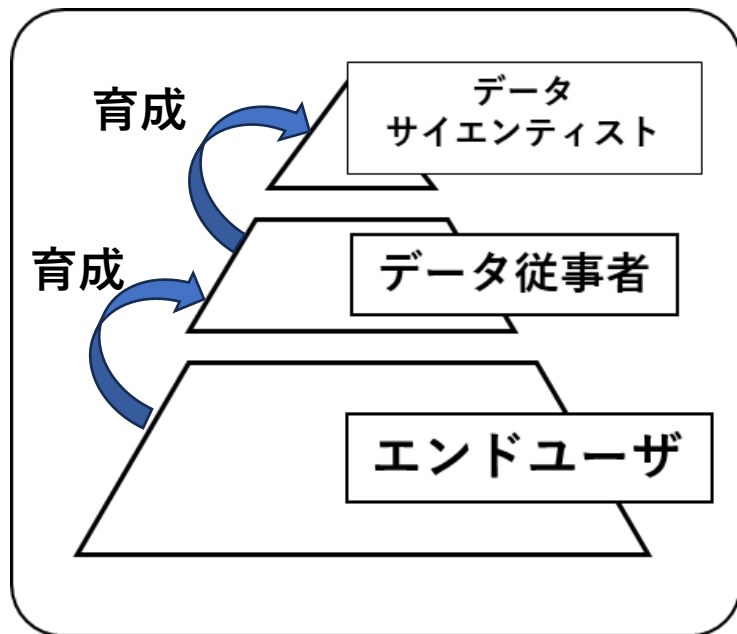
生成AI以前(2016年)

生成AI以後(2022年~)

統計学
統計ソフト

業務システム
表計算
DB

スマホ
PC



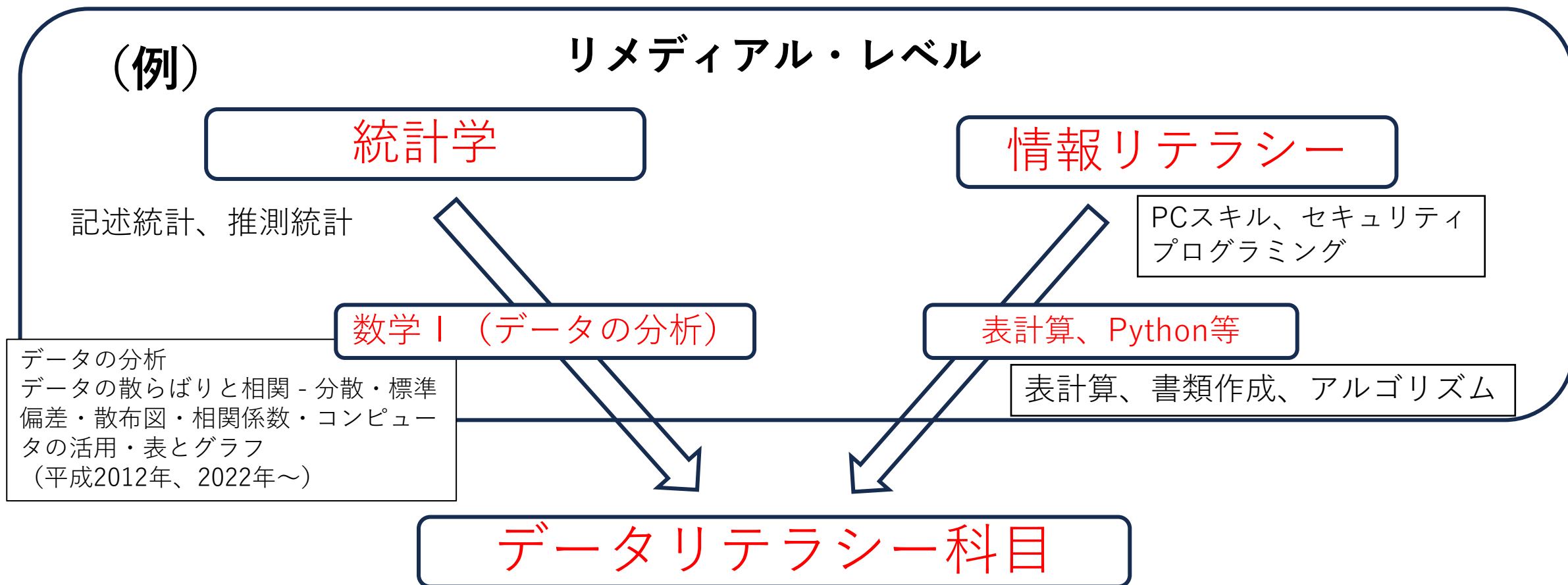
例) ExcelはMicrosoft 365 Copilotによる生成AIの活用

リテラシーレベル モデルカリキュラムの構成

- モデルカリキュラムの構成を以下のとおり「導入」「基礎」「心得」「選択」に分類し、学修項目を体系的に示した。
- 「導入」「基礎」「心得」はコア学修項目として位置付ける。「選択」は学生の学習歴や習熟度合い等に応じて、適切に選択頂くことを想定している
- 次頁よりそれぞれの分類における「学修目標」「学修内容」「スキルセット（キーワード）」をまとめた。

導入	1. 社会におけるデータ・AI利活用 1-1. 社会で起きている変化 1-2. 社会で活用されているデータ 1-3. データ・AIの活用領域 1-4. データ・AI利活用のための技術 1-5. データ・AI利活用の現場 1-6. データ・AI利活用の最新動向
基礎	2. データリテラシー 2-1. データを読む 2-2. データを説明する 2-3. データを扱う
心得	3. データ・AI利活用における留意事項 3-1. データ・AIを扱う上での留意事項 3-2. データを守る上での留意事項
選択	4. オプション 4-1. 統計および数理基礎 4-2. アルゴリズム基礎 4-3. データ構造とプログラミング基礎 4-4. 時系列データ解析 4-5. テキスト解析 4-6. 画像解析 4-7. データハンドリング 4-8. データ活用実践（教師あり学習） 4-9. データ活用実践（教師なし学習）

共通科目におけるデータリテラシーの位置づけ



データを読む、データを説明する、データを扱う

共通教育のデータリテラシー教育の課題

- Python
 - 文系で扱うのは時間がかかる（PC嫌い）
- 統計ソフト：
 - 有料ソフト
 - JMP, SPSS, SAS, Excel等
 - フリーソフト
 - JASP、Rや環境構築（個人PC）などへの導入
- 自分の授業ではExcelの採用
 - MSオフィスの普及
 - データと分析の両方を同時に学習
 - Excelの操作の授業になっている⇒データ分析をおざなり

既存のデータ分析学習手順

1. 分析手法に合ったシナリオ
2. 分析用の仮想データ提示（Excelファイル）
3. 分析方法の紹介
4. 分析実践

Excel等によるデータ分析の実践

⇒ここをChatGPT 置き換えると・・・？

生成AIによるデータ分析の試み

- データリテラシーの演習内容を**Excelによる分析からChatGPTによるデータ分析に変更すると？**

1. 学習する分析方法をデータ分析をChatGPTでやってみた。

⇒ファイル名に分析方法をつける。

「ヒストグラム_A組B組.xlsx」

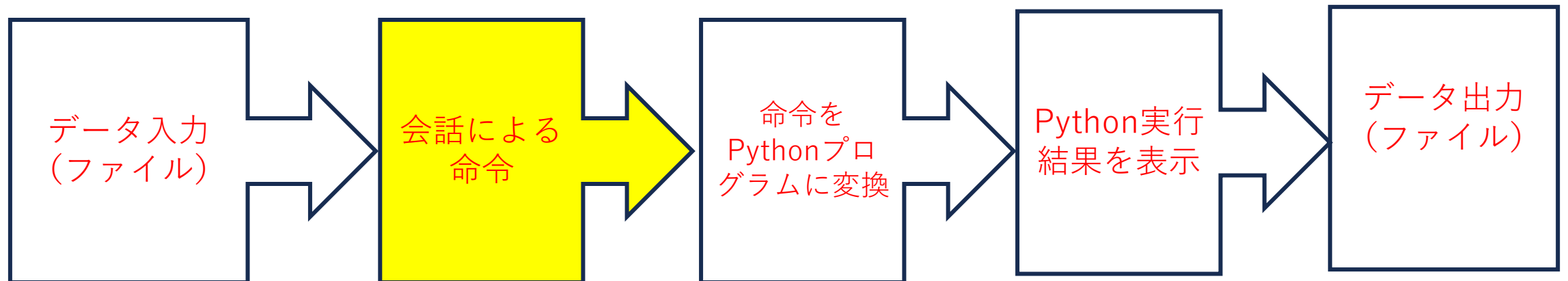
2. 分析方法の指示なしでデータ分析をやってみた。

⇒ファイル名は「data00.xlsx」

ChatGPT4 (Advanced Data Analysis : ADA)

- Advanced Data Analysis(元Code Interpreter)

- ファイルをアップロードすることができる。
- ファイルのデータを使って生成AIにより命令できる。
 - 命令を解釈して、その目的に合ったPythonを実行した結果を表示する
- 完成したファイルをダウンロードもできる。



深呼吸して、Step-by-stepで作成してください

Advanced Data Analysisで可能なこと

• データ解析

- オーディオファイルの解析
- シミュレーションやモデリング
- プログラミング学習
- コーディングサポート
- 文書の解析
- ファイル形成、変換
- QRコード生成
- 画像処理

• 長期記憶

- チャットでより多くのやり取りで分析できる

• 作成可能なグラフ

- 折れ線グラフ/エリアグラフ
- 棒グラフ/ヒストグラム (+エラーバー)
- 円グラフ
- 散布図
- 箱ひげ図
- ペアプロット
- ヒートマップ
- **3D**グラフ
- 等高線図

結論（教育プロセスの変化）

- 分析手法スキルの習得⇒生成されるデータ分析からの統計的アプローチを習得

既存の“教育方法”

1. データ分析のシナリオを説明する
2. 分析したいデータを確認する
3. 学習する分析方法を説明する
4. データ加工
5. 統計ソフトによる分析

生成AIを使用した“教育方法”

1. 分析したいデータを目の前にする
2. データを生成AIに読み込ませる
3. 提案されるデータ分析を検討する
(生成AIから分析の流れやシナリオが提案される)
4. 結果を検証する
 - 分析結果を読む
 - **知らない分析方法は追って習得する)**

期待

- 気軽にデータ分析しようという動機（ハードルが下がる）
- アクティブラーニングやプロジェクト学習におけるデータ取得への動機

ChatGPTによるデータ分析

- ChatGPTにより以下の処理が生成可能となった
 - データ加工
 - データ分析
 - グラフ
 - レポート
- 解釈にも統計の知識の必要
 - データから求めたいものを定める（目的の設定）
 - 分析方法の知識（関数など覚えなくてよい）⇒分析結果の読み取り力の向上（統計レポートの解読）

直近の課題

(生成AI、授業スタイル、対応策)

• 分析結果などに揺らぎの影響がある

- 実行するごとに結果が変わる
 - 例) たまに分析途中で下1桁が0に丸められるなど
- Open APIの使用を検討? (ただし、APIでADAを実現させるのは難しい)

• 有料への対応

- 大学組織でのChatGPT Enterpriseの検討

• 情報漏洩への危惧

- 入力したデータがAIのための学習データに用いられる可能性
- △ ChatGPT Plus (個人有償プラン)
 - Advanced Data AnalysisやPluginを用いる際、「Chat history & training」をオフにできない
- ○ ChatGPT Enterprise プラン
 - 「Chat history & training」をオフにできる

データサイエンスにおける 学習者の生成AIへのアプローチ

微修正（自分）



目指す答えに向けて、プロンプトを試行する

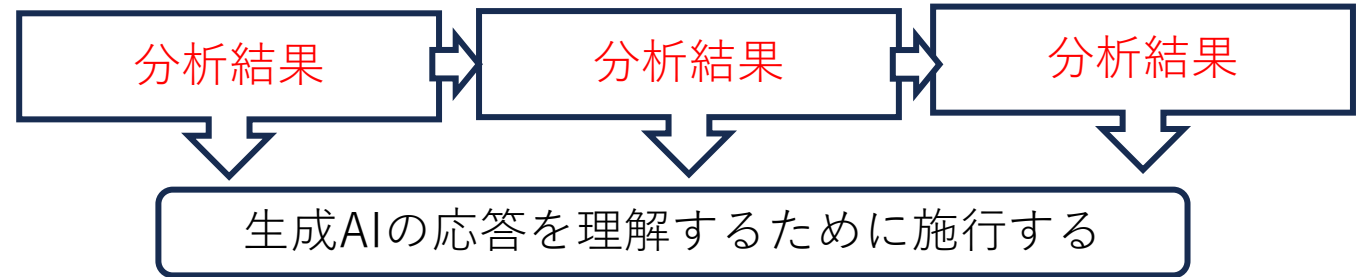
省力化
コスト削減



応答結果のイメージがある

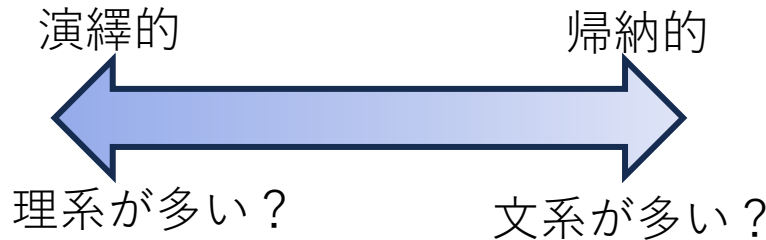
何から始めたらいいのかわからない人のハードルを下げる

理解するために複数回問い合わせする



いくつかの結果を総合して結果について考える

とりあえずAIに相談する



作文とかなら逆になる…かも



応答結果のイメージはない

教える側の変化

- 「データ加工やデータ分析の紹介」から「データリテラシー」へ
 - データ加工やExcelの学習配分を減らす。
 - 生成AIが分析方法の提案やデータ処理を行うので、データの分析結果に対して学習を行うことができる。
 - 分析に対して懐疑的にアプローチさせる経験も。。。！？
- プロジェクト型、アクティブラーニング型への移行
 - 企業データの活用や社会調査を使ったプロジェクトとの連携
 - 分析結果の検討や考察に時間をかけることができる。
 - 実データの分析など分析可能かどうかわからないデータに対しても対応できる。

学ぶ側（学生）の変化

- 分析のためのスキルなしで学べる
 - Excelの使用方法やデータ加工，分析のためのスキルなど直接求められない
- 生成AIが生成する結果に対して試行錯誤や必要な知識の学習が求められる
 - 生成AIの分析に習得していない分析方法の提案や結果が提示される
 - 的確に指示する技術、言葉選び
 - プロンプトエンジニアリング
 - 統計用語の活用方法
- 再現性、揺らぎ対策
 - プロンプト履歴のログ管理の技術
- 学生が懐疑的に見ることができるとのか
 - 処理の内容を知るためにPythonの知識が必要となる

新たな課題

- AIセキュリティ対策
- 生成AIによるハルシネーション対策
 - 生成AIが出すもっともらしい分析結果に対して、学生が懐疑的に見ることができるとか、判断等できるようになるのか
 - **クリティカル・シンキング** ⇒ 本来のデータリテラシー
 - 元データのデータ加工の過程や用いられた分析方法に関する知識
 - 正確に分析させるためにプロンプトエンジニアリング
 - Python, 他の分析ツールの活用の必要性
 - 提示されたデータの審議を確認

⇒ **(結局、統計学、統計ツールやPythonなどを学ぶ必要があるのでは…)**