

2023年10月13日



中級プログラミングによる 生成AIの活用事例紹介

伊東 栄典

九州大学 情報基盤研究開発センター

ito.eisuke.532@m.kyushu-u.ac.jp

Who am I?

情報基盤運営と教育

- 九州大学情報基盤研究開発センター・准教授
 - 電子認証基盤, 電子メール, 情報共有
- 講義
 - プログラミング (C言語, Python)
 - コンピュータ・システム
 - オペレーションズ・リサーチ
 - オペレーティングシステム
 - 人工知能
 - 暗号と情報セキュリティ

研究テーマ

データサイエンス系

- 大規模文書群からの知識発見
- 最近の研究
 - ネットコンテンツ分析
 - マンガ画像分析
 - スマホ学生証

本講演で伝えたいこと

- **文章生成AIは、プログラミングを効率化**
 - コードの書き方が分からない場合
 - 経験の有る人も、少ない人も効率化
- **ただしプログラムを理解できないとだめ**
 - 自分が求めるプログラムかどうかの判断
- **使ってきた感想**
 - **初級・中級の課題は、答えに辿り着きやすい**
 - **難しい課題の場合、答えに辿り着きにくい**
- **プロンプトエンジニアリングの技能が必要**
 - プロンプト文の書き方
 - Chain of Thought と呼ぶ試行錯誤が必要

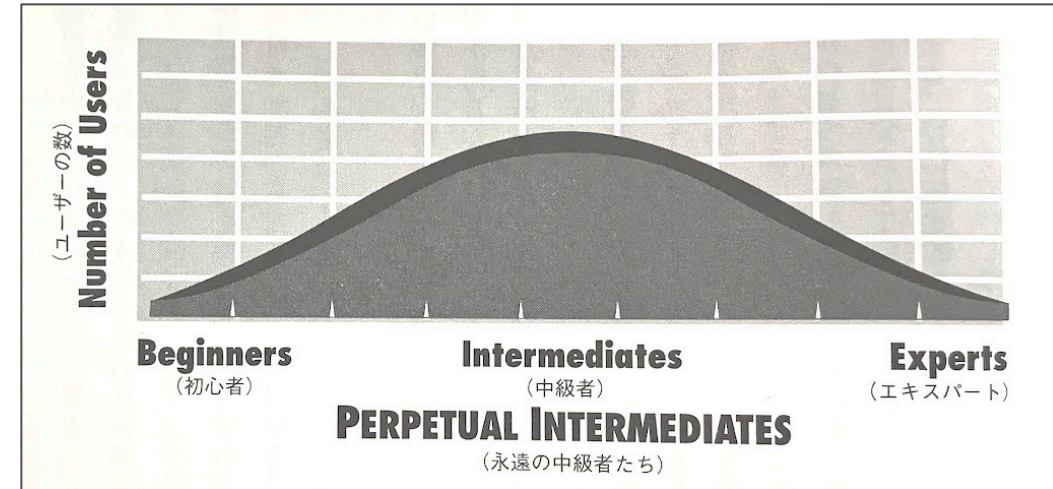
1. はじめに

1. はじめに
2. 初級の課題
3. 中級の課題
4. 上級の課題
5. おわりに

1. はじめに

- 生成 AI で色々な悩みが解決
 - 自然言語で問いかけるだけで解決できる事が多い。
- 本講演に対する自問自答
 - AI技術を使うものの、生成AIを作る専門家ではない。
 - 生成AIの使い方も、普通のレベル（だろう）。
 - 多くの方は「**永遠の中級者**」である。
 - **中級者の活用事例紹介**には意義があるだろう。
- 何に活用しているか
 - **プログラム作成, 解説**
 - 情報ツールの使い方
 - マインドマップ・シーケンス図の作成
 - 文章の作成・要約・添削・翻訳

永遠の中級者



多くの初心者は中級者にステップアップするが、ほとんどの中級者は上級者にステップアップしない。

Alan Cooper 著, 山形 浩生 訳, コンピュータはむずかしすぎて使えない, p.375, 翔泳社, 2000.

プログラムの作成・解説

- 「〇〇言語で、△を実現するプログラムを教えて」と聞くと、**コード片を出力**
 - Python, C, HTMLなどのコードを出力。
 - プログラムを**書ける人**には**超便利**
- 何が嬉しいか？
 - 忘れていた処理
 - あまり経験の無い処理
 - 面倒な定形処理
 - 誕生日入力や住所入力など
- プログラムのコードを読ませ、解説を頼むと、わりと良い説明をしてくれる

利用してきたプログラミング言語

言語	時期	説明
C		現在は授業でのみ利用
Perl	-- 2000	研究や業務に活用
Ruby	2000 -- 2015	研究や業務に活用 主にデータ分析
Python	2015 --	研究や業務に活用 データ分析や機械学習など

活用事例を、難易度が
初級・中級・上級の課題
について紹介。
(主観的な難易度)

2. 初級の課題

1. はじめに
2. 初級の課題
3. 中級の課題
4. 上級の課題
5. おわりに

2. 初級の課題

- 簡単な課題を文章生成AIに尋ねると、だいたい回答してくれる。
 - クイズ的な質問
 - 簡単な文章作成
 - 簡単な分類
 - 簡単なプログラムの作成と解説
- 大学2年生向けの「**c言語プログラミング演習**」を担当
 - 課題が20個以上ある。
 - 生成AIに尋ねると、**だいたい解けてしまう**。
- 生徒には
 - 「禁止しないけれど、自分で作らないと能力が伸びない」と説明。
 - また「ChatGPTの癖・言い回し（コードやテキスト）が分かる」と伝える。

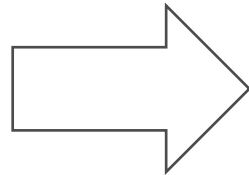
課題 scanf の使い方

プログラムを入力して以下のことを確かめよ。

- int の範囲を超える数を入力するとどうなるか
- 小数を入力するとどうなるか
- 文字を入力するとどうなるか

```
#include <stdio.h>
main()
{
    int i;

    printf("i ? ");
    scanf("%d",&i);
    printf("i = %d\n",i);
}
```



提供されたプログラムは、整数値を入力してそれを表示するシンプルなCプログラムです。プログラムを実行すると、異なる入力に対する動作が以下ようになります。

1. `int` の範囲を超える数を入力するとどうなるか :

- `int` 型は通常、特定の範囲内の整数値を表現できます。範囲を超える大きな数を入力すると、オーバーフローが発生し、値が不正確になります。例えば、**2147483647** を超える値を入力すると、結果が予測不能になる可能性があります。

2. 小数を入力するとどうなるか :

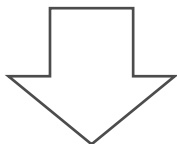
- `scanf` 関数は `%d` フォーマット指定子を使用して整数を読み取るため、小数点以下の値を入力すると、**小数点以下の部分は切り捨て**られます。つまり、小数部分は無視され、整数部分だけが変数 `i` に格納されます。

3. 文字を入力するとどうなるか :

- `scanf` 関数は `%d` フォーマット指定子を使用して整数を読み取るため、文字を入力すると変換エラーが発生します。プログラムが整数を期待しているため、文字を入力するとエラーメッセージが表示され、`i` の値は変更されません。

課題 素数の算出

エラトステネスの篩の方法を用いて、1000 以下の素数を入力するC言語プログラムを作成してください。



以下は、エラトステネスの篩の方法を用いて 1000 以下の素数を入力する C 言語プログラムの例です。

このプログラムでは、まず配列`is_prime`を用意し、最初にすべての要素を`true`に初期化します。次に、2 から \sqrt{N} までのすべての数について、もし`is_prime[i]`が`true`であれば、 i の倍数の要素をすべて`false`に変更します。最後に、`is_prime`配列の中で`true`の要素を持つインデックスを出力します。

```
#include <stdio.h>
#include <stdbool.h>
#define N 1000

int main() {
    bool is_prime[N+1]; // 篩の状態を保存する配列
    int i, j;

    // 初期化
    for (i = 2; i <= N; i++) {
        is_prime[i] = true;
    }

    // 篩をかける
    for (i = 2; i * i <= N; i++) {
        if (is_prime[i]) {
            for (j = i * i; j <= N; j += i) {
                is_prime[j] = false;
            }
        }
    }

    // 結果を出力
    for (i = 2; i <= N; i++) {
        if (is_prime[i]) {
            printf("%d ", i);
        }
    }
    printf("\n");
    return 0;
}
```

初級の課題のまとめ

- 自分で使う場合は便利
 - 検索エンジンより生成AI
- プログラミングの課題
 - 生成AIを前提とした課題を作るべきなのか？
 - それはどのような課題なのか？

3. 中級の課題

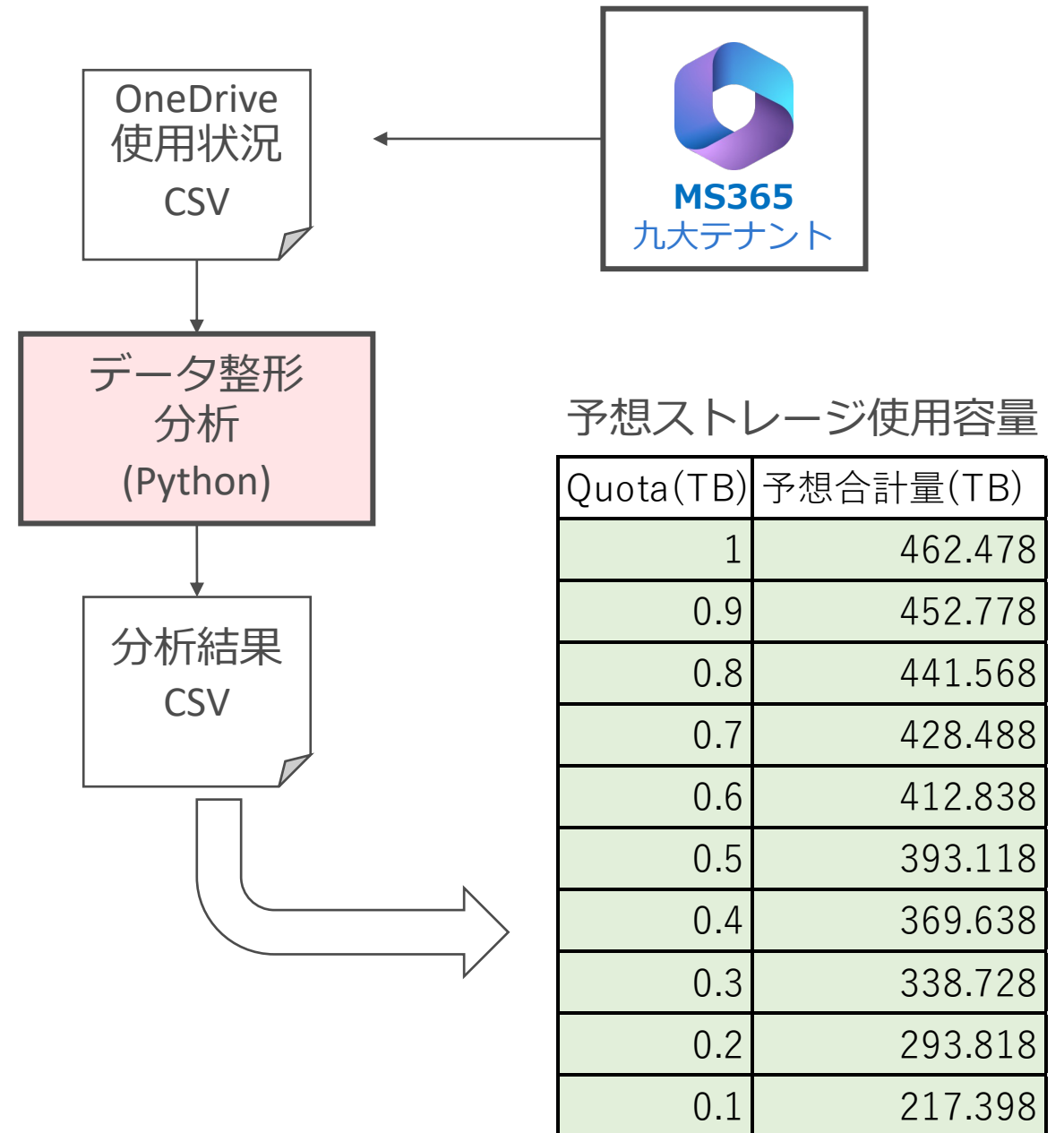
1. はじめに
2. 初級の課題
3. 中級の課題
4. 上級の課題
5. おわりに

3. 中級の課題

- ある程度の専門知識が必要な課題について，生成AIに支援してもらう。
- 2つの事例
 - 大学の情報センターでのデータ分析プログラム作成
 - システム管理：サーバ証明書の作成方法
- 他にも
 - サーバ機器の各種設定方法
 - 論文の校正

データ分析プログラム作成

- MS365のストレージ利用状況を調査。
- データ (CSVファイル) を, Pythonで整形, 分析。
- 結果
 - 全学で535.7TB利用
 - 1TB以上の方は91人で, 164.23TB使用 (30.66%)
- データの分析結果から, 人手で使用容量予測を算出。
- 以前は全部自分で書いていた
- 生成AIに自然言語で質問することで, 大部分が書けてしまう。



ストレージ使用量分析のコード作成

python言語の jupyter notebook を書いて、**CSVファイルのデータ**分析を行います。データは、大学が提供する共有ストレージの、各学生ごとの使用量です。CSVファイルは、一行目が項目名になっていて、それは「'date', 'URL', 'name', 'daleted', 'l_date', 'files', 'files_a', 'byte', 'quota', 'mail', 'period'」です。この**データを、pandas に読み込む方法**を教えてください。

作成された df から、**name, byte, mail の列だけからなる新しいdf**を作りたい。その方法を書いてください。

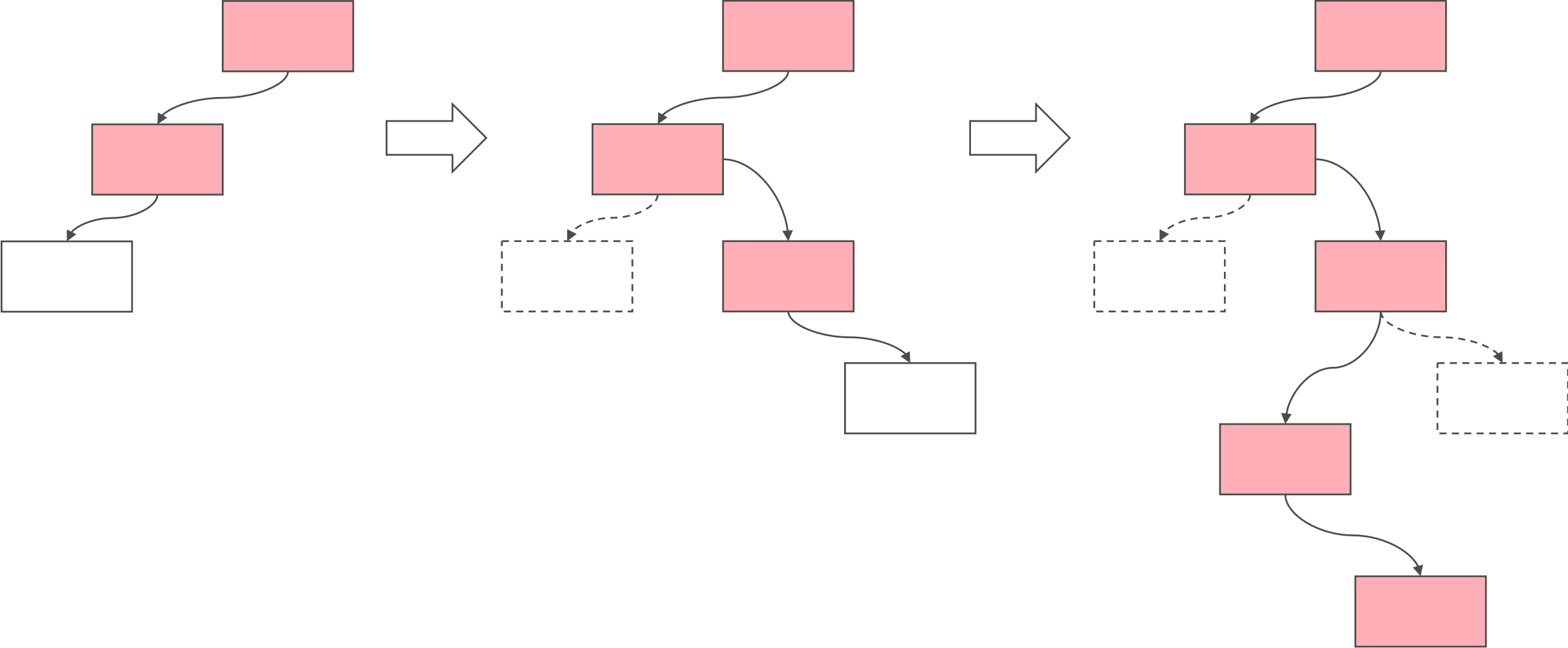
メールアドレスの@以降が、学生は「s.kyushu-u.ac.jp」で、職員は「m.kyushu-u.ac.jp」です。mail列の文字列から、dfを**学生だけのs_df**と、**職員だけのm_df**に分ける方法を教えてください。

dfのデータを、**byteの数値が大きい順に並べ替える**方法を書いてください。

dfのデータで、byteの値が1TB以上の人だけに限定して**byteの値を足し合わせて**、合計ストレージ使用量と、その人数を算出してください。

最後に、今までの処理を**まとめて1つのプログラム**にしてください。

ストレージ使用量分析のコード作成



サーバ証明書の秘密鍵・公開鍵作成

- 公的なWebサーバを運用する場合, サーバ証明書を使う
 - サーバの真正性確認と, 通信の暗号化
- NII・学認ではサーバ証明書発行サービスを提供
- サーバ証明書発行のためのマニュアルが学内サイトにある
 - コマンドを打てば必要データを得られるけれど, その意味は書いていない。
 - 意味を知りたい場合, 生成AIに聞くと回答が得られる。

以下のコマンドがどんな動作をするのかを教えてください。

```
openssl genrsa -des3 -rand file1.txt:file2.txt:file3.txt 2048 > servername.key
```

このコマンドは、OpenSSLを使用してRSA秘密鍵を生成し、それをDES3暗号で保護し、指定した乱数ファイルを使用して鍵の生成に影響を与えます。生成された秘密鍵は servername.key という名前のファイルに保存されます。以下はコマンドの各部分の説明です：

- openssl genrsa: OpenSSLのコマンドで、**RSA暗号の秘密鍵を生成**するために使用されます。
- -des3: このオプションは、生成された**秘密鍵をTriple DES (DES3) 暗号で保護**することを指示します。これにより、秘密鍵ファイルがパスワードで保護され、読み取り時にパスワードが必要になります。
- -rand file1.txt:file2.txt:file3.txt: このオプションは、秘密鍵の生成にランダム性を導入するために指定された**ファイルを乱数源として使用**します。指定されたファイルの内容が秘密鍵生成に影響を与えます。
- 2048: 生成される**RSA秘密鍵のビット数を指定**します。この場合、2048ビットの鍵が生成されます。一般的に、鍵の長さが長いほどセキュリティが向上しますが、同時に計算コストも増加します。
- > servername.key: 生成された秘密鍵は、servername.key という名前のファイルにリダイレクトされます。つまり、このファイルに保存されます。

最終的に、このコマンドを実行すると、Triple DESで保護された2048ビットのRSA秘密鍵が生成され、servername.key ファイルに保存されます。この秘密鍵を使用して、SSL/TLS通信を設定するなど、さまざまなセキュアな通信プロトコルをサポートするために利用できます。秘密鍵を使用する際には、指定した -des3 オプションで設定したパスワードが必要になります。

中級の課題のまとめ

- 自分で使う場合
 - かなりの時短効果
 - 自分が知っているため、質問文が良く、回答も理解できる
 - 試行錯誤の回数が少ない
- 研究室の学生
 - 「○○についてはChatGPTで尋ねてみて」でかなりの部分が解決
 - 学生の方が、使い方が上手

4. 上級の課題

1. はじめに
2. 初級の課題
3. 中級の課題
4. 上級の課題
5. おわりに

4. 上級の課題

- 難易度の高い課題をプログラムで処理したい
- 2つの事例
 - デジタル学生証（ワクチン接種証明）の処理
 - 回帰分析の例

QRコードとデジタル学生証

- ワクチン接種証明のように，スマートフォンで学生証が作りたい。
- QRコード，その内容（テキスト）など，調べることは多い。
- 調べた後，プログラムを作る必要がある。

こんなのが作りたい



ワクチン接種証明



日本のデジタル庁が **Covid19 ワクチン接種証明のアプリ** を提供しています。ワクチン接種済の場合、その証明書をスマートフォンに格納し、接種証明書を他者へQRコードとして提示します。この **QRコードの内容** について教えてください。

申し訳ございませんが、私の知識は2021年9月までのものであり、日本のデジタル庁が具体的なCovid-19ワクチン接種証明のアプリやそのQRコードの内容についての**情報は含まれていません**。Covid-19関連のアプリやサービスは急速に変化しており、新しい情報やアップデートがあるかもしれません。

Covid-19ワクチン接種証明のアプリやQRコードに関する正確な情報を入手するためには、日本のデジタル庁の公式ウェブサイトや関連する公式発表を確認することをお勧めします。また、QRコードの内容はプライバシーやセキュリティの観点からも重要ですので、公式情報源から提供される情報に従うことが重要です。

ChatGPT 3.5 は回答できなかった。

日本のデジタル庁が **Covid19 ワクチン接種証明のアプリ** を提供しています。ワクチン接種済の場合、その証明書をスマートフォンに格納し、接種証明書を他者へQRコードとして提示します。この**QRコードの内容**について教えてください。

ワクチン接種証明のQRコードの情報は、**SMART Health Cards (SHC)** というデータ形式で格納されています。SHCは、ワクチン接種やPCR検査などの健康情報の証明書を電子的に保存するための国際標準規格です。

SHCのQRコードには、以下の情報が含まれます。

ヘッダー：QRコードの種類やバージョンなどの情報が含まれます。

データ：氏名、生年月日、接種記録、接種済証の種類、接種済証の有効期限などの情報が含まれます。

電子署名：証明書のデータが偽造されていないことを確認するための情報です。

ヘッダーには、QRコードの種類やバージョンなどの情報が含まれます。データには、氏名、生年月日、接種記録、接種済証の種類、接種済証の有効期限などの情報が含まれます。電子署名には、証明書のデータが偽造されていないことを確認するための情報が含まれます。

：

：

QRコード画像の処理など

QRコード画像を解読するPythonコードを書いてください。

zbarやpyzbarとは何ですか？

→ 画像処理のライブラリ

MacOS で zbar を入れましたが、pyzbar が動きません。なぜでしょうか。

→ バージョン間違いで動かなかっただけだった。

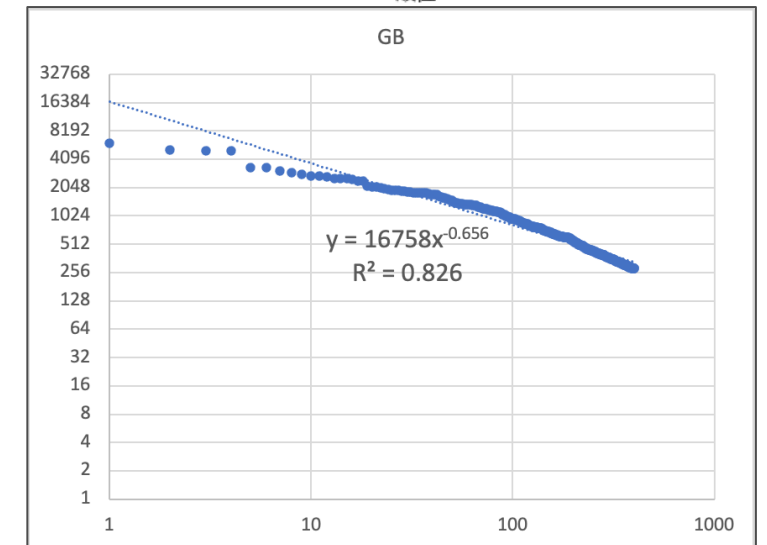
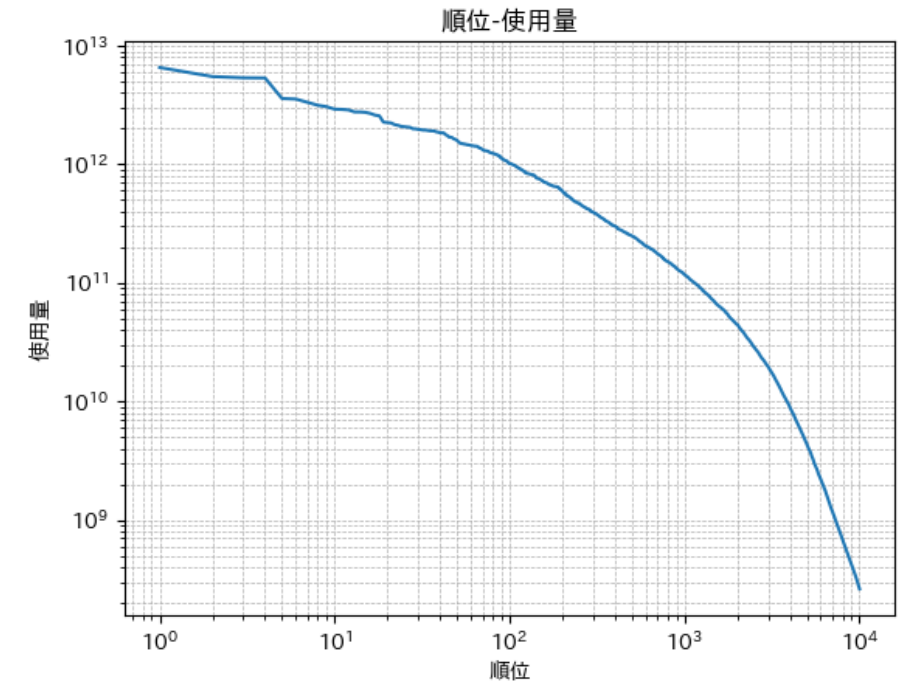
JWTとは何ですか？

→ JSON Web Token の略で、...

- W3C の Verifiable Credential 規格資料を**事前に勉強**していたため、**内容を理解**できている。
- そもそも生成AIが学習する文書が少ない（ように思われる）。
- 自分の質問文が**求める内容を文章化できていない**のかも。
- 生成AIの時短効果は少ない。

回帰分析の例

- MS365のストレージ利用状況。
- 対数正規分布になると知られている。
- 回帰分析でパラメータを推定。
- ChatGPTに「ストレージ」という単語をいれると、回答が引っ張られる。
- 欲しい近似曲線のパラメータが得るまでの作業が膨大になる。
 - 試行錯誤が増える
 - 専門用語やコードの調査学習



上級の課題のまとめ

- 生成AIの時短効果は少ない。
 - プロンプト文を作れていない：その分野の専門用語を知らない
- 事前に勉強しておかないと，AIの文章が理解できない。
 - W3C の Verifiable Credential 規格資料
 - 回帰分析の知識
 - 生成されたコードの処理内容
- そもそも生成AIが学習に使う文書が少ない（ように思われる）
- **結局，従来からの調査研究と同様になる。**
 - AIを含む様々な資料やツールを用いての調査
 - 試行錯誤
 - 内容理解

6. おわりに

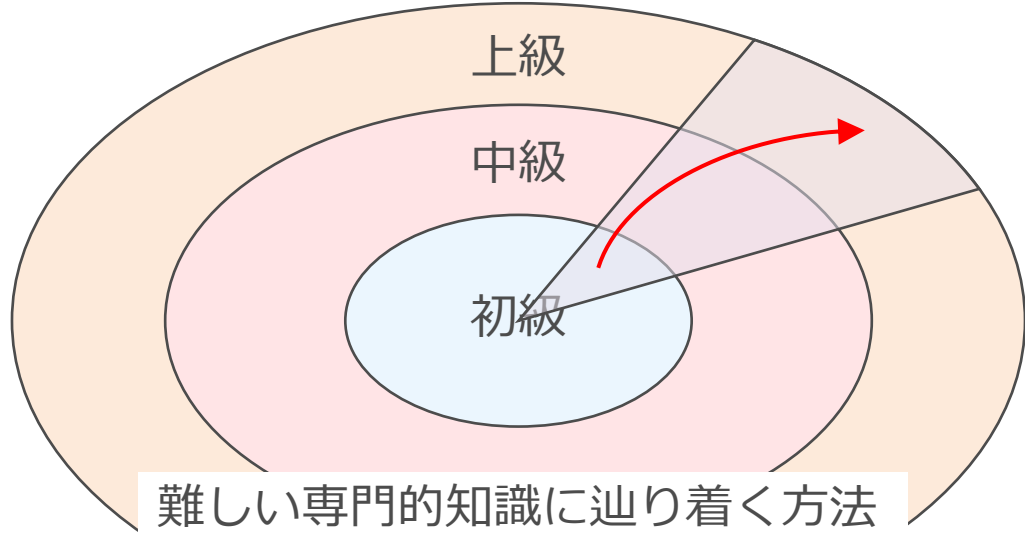
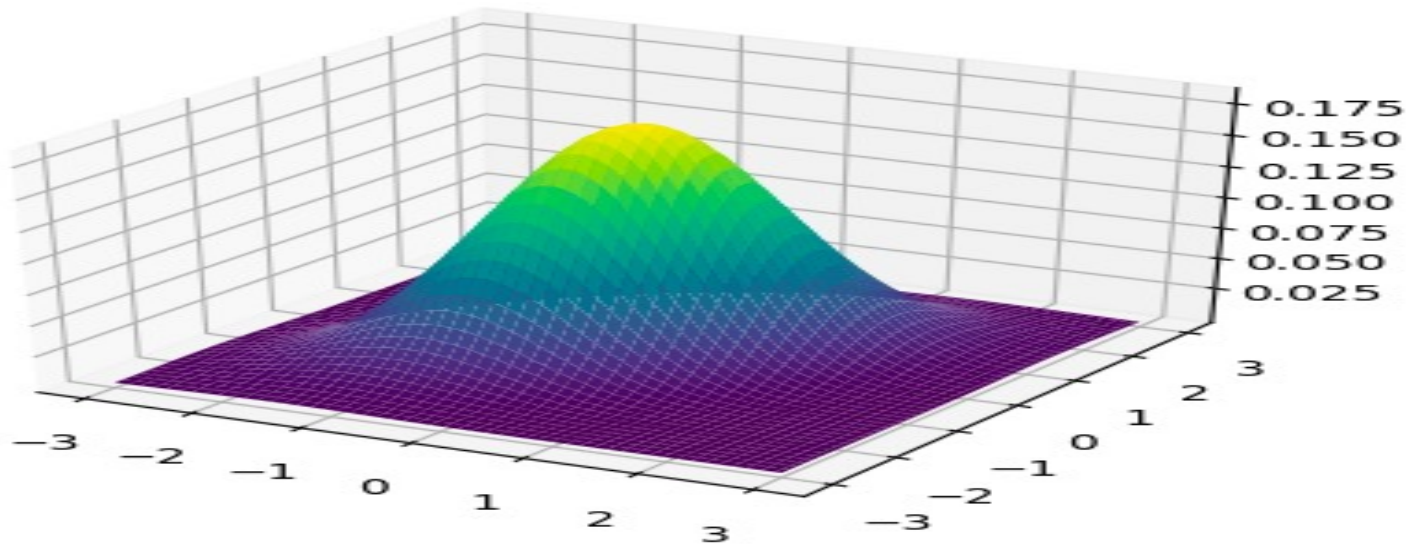
1. はじめに
2. 初級の課題
3. 中級の課題
4. 上級の課題
5. おわりに

6. おわりに（本講演で伝えたいこと）

- **文章生成AIは、プログラミングを効率化**
 - コードの書き方が分からない場合
 - 経験の有る人も、少ない人も効率化
- **ただしプログラムを理解できないとだめ**
 - 自分が求めるプログラムかどうかの判断
- **使ってきた感想**
 - **初級・中級の課題は、答えに辿り着きやすい**
 - **難しい課題の場合、答えに辿り着きにくい**
- **プロンプトエンジニアリングの技能が必要**
 - プロンプト文の書き方
 - Chain of Thought と呼ぶ試行錯誤が必要



ゴルフのティーショット。グリーンが複数あるため、どの方向に打てばよいか迷うゴルファーの図 (MS Bing Image Creator)



難しい専門的知識に辿り着く方法
 • 文章生成AIは専門知識が苦手？

専門家・研究機関は必要



ゴルフのティーショット。グリーンが複数あるため、どの方向に打てばよいか迷うゴルファーの図 (MS Bing Image Creator)