

第一回 OOL Tech Connect

AI × DataOps

- AIを活用したデータ利活用促進の取り組み -

TIS株式会社

デジタル社会サービス企画ユニット

デジタル社会サービス企画部

坂本 諒太

- 自己紹介
- OOLでの取り組み
- DataOpsについて
- DataOps PJで検討しているAI × DataOps
 - 本日、紹介する検証テーマ
- 検証 1. 計画立案
- 検証 4. データの処理
- 検証 6. フィードバック/現状把握
- DataOps PJ 今後の予定
- まとめ



プロフィール

坂本 諒太
Ryota Sakamoto

TIS株式会社
デジタル社会サービス企画部

趣味など
OSS開発、テニス

SNS

Twitter
@SkeinIn

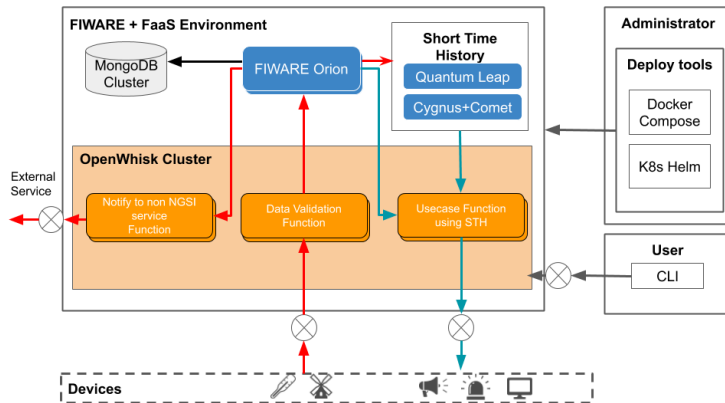
Github
@ryo-ma

略歴

- ・ TIS株式会社に入社後、研究開発部門にてクラウド自動化の研究開発およびそれに伴ったOSSの開発に従事
- ・ 大阪大学に常駐し対話/社会心理学の研究に携わり、その知見を活用した高齢者向け生活支援AI対話サービスをスタートアップエンジニアとしてスタートアップ立ち上げに従事
- ・ FIWAREによるIoT/ロボティクスプラットフォームの研究開発に従事し、会津若松にて公道走行配送ロボットの実証実験を担当
- ・ 現在はデータ利活用という観点で社会のあるべき姿を検討し社会実装に向けた活動を実施

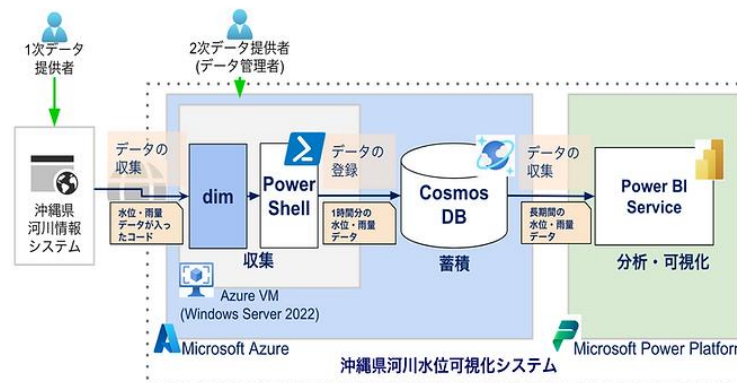
2019~ IoTプラットフォームPJ

- FIWARE 検証
- FIWARE Meteoroid開発



2021~ DataOps PJ

- DataOpsプロセス検討
- DataOps基盤検証

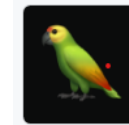


現在

2023~ DataOps PJ

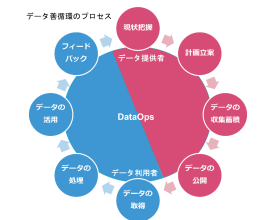
- AI × DataOps検証
- ユースケース検証

OpenAI



LangChain

220 followers



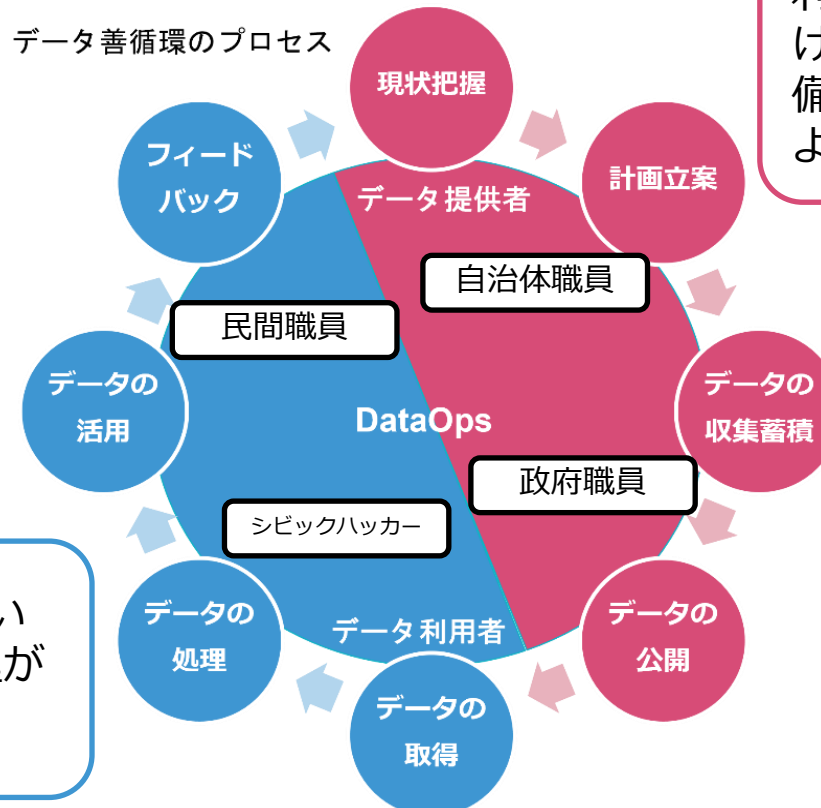
- DataOpsとは「利用者がデータを欲しい時に素早く信頼できる形で取得できるようにデータ利用者と管理者とが協調してプロセス、ツール環境や組織文化を継続的に改善する取り組み」

出典, Gartner : <https://www.gartner.com/en/newsroom/press-releases/2018-09-11-gartner-hype-cycle-for-data-management-positions-three-technologies-in-the-innovation-trigger-phase-in-2018>

オープンデータにフォーカス オープンデータが利活用されない課題に対してアプローチ

TISが検討している オープンデータを想定した DataOpsのプロセス図

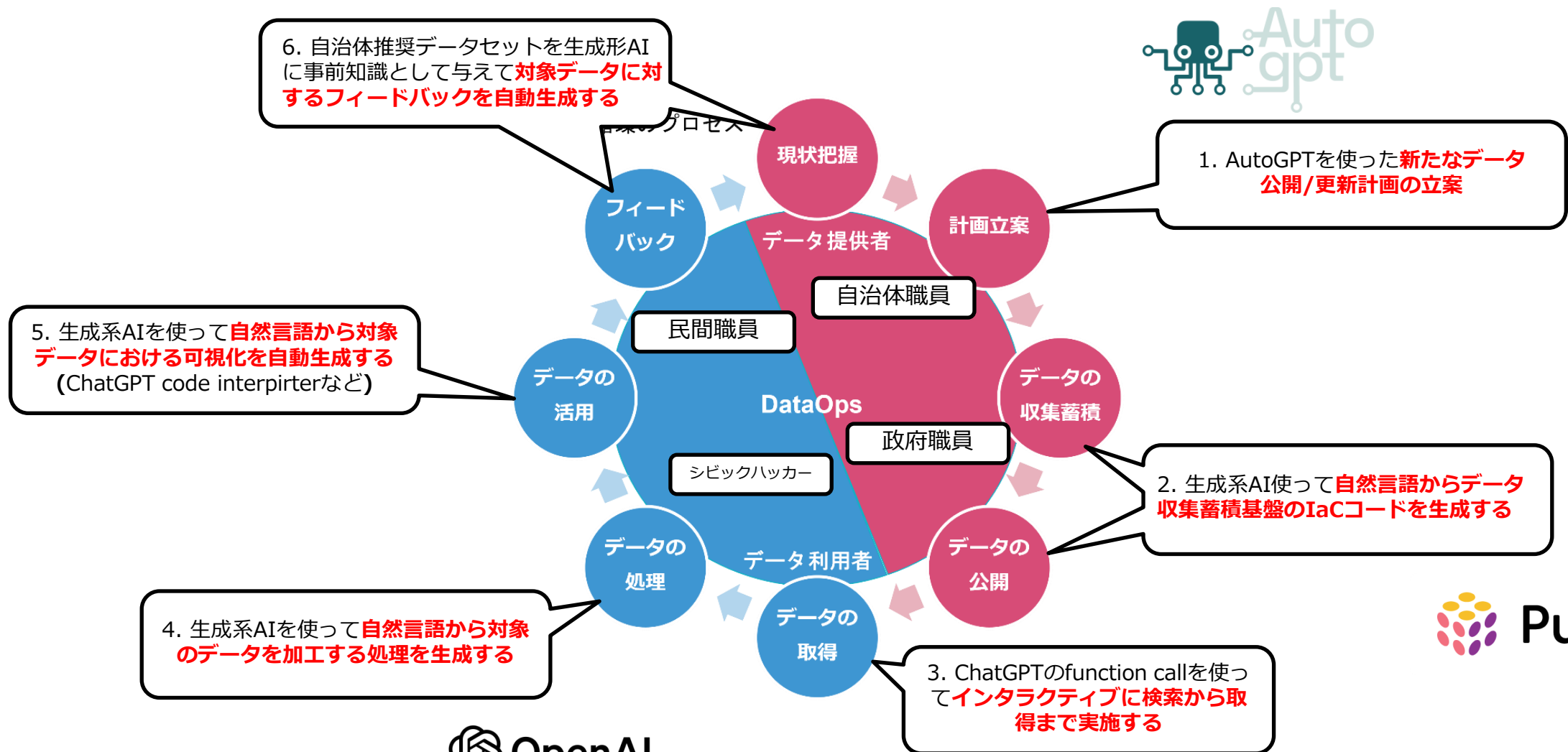
データ善循環のプロセス



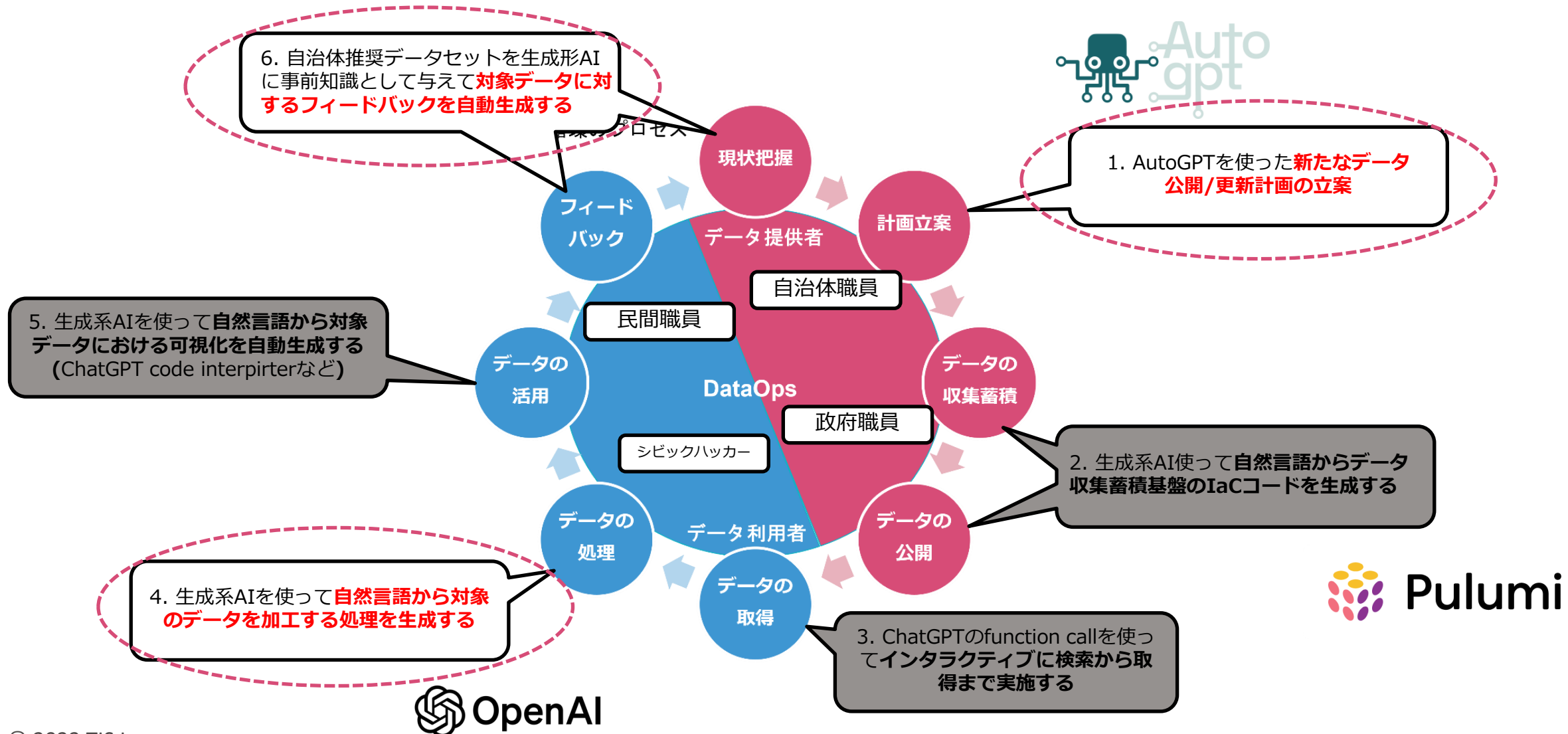
利用者からのフィードバックを受けてニーズを考慮したデータの整備/公開を実現することで利用者による利活用につながる

提供者から利用者のニーズが高いデータやデータ形式で取得/処理ができ利活用事例が促進される

昨今のAI技術(生成系AI)の活用によりDataOpsのプロセス改善や新たな価値向上の可能性を模索



本日は**検証1. 検証4. 検証6.**を紹介



■ DataOpsプロセス

□ 計画立案

■ 使用技術

□ Auto-GPT

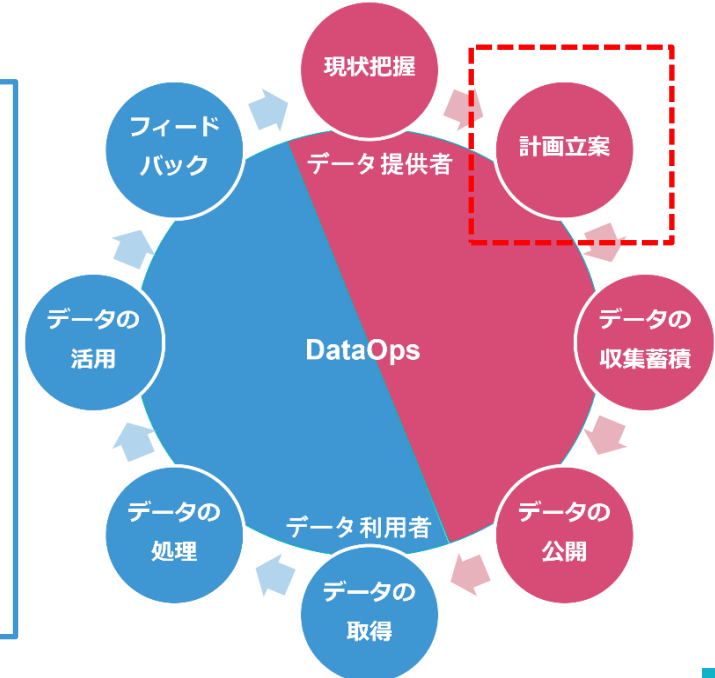
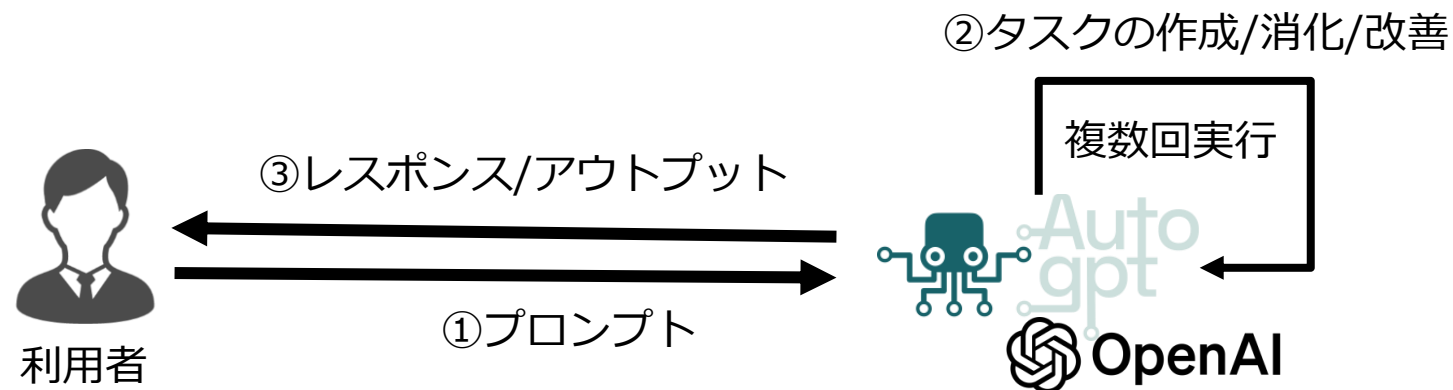
- ユーザは目標設定を行うと、自動でそれに必要なタスクを複数作成し、タスクの消化、改善まで自動で実施してくれるOSS
- 内部的にはOpenAIのGPTモデルを使用している
- <https://github.com/Significant-Gravitas/Auto-GPT>



■ 検証

□ 1. AutoGPTを使った新たなデータ公開/更新計画の立案

検証イメージ



検証 1. ユーザプロンプトからAutoGPTによる実行の流れ



AutoGPTに与えたプロンプト

日本の自治体内でオープンデータの推進活動始めるために、技術を知らない上司を説得する資料を作成しろ



AutoGPTがゴールを作成 & タスクを計画しながら実行

- オープンデータ推進活動の**メリット**や**重要性**を、技術者ではない自治体の上司に**効果的に伝える説得力**のある文書を作成する。
- オープンデータの背景にある技術について、**非技術者にも理解しやすいよう**に明確かつ簡潔に説明する。
- 地方自治体の文脈の中で、オープンデータの取り組みが透明性、市民参加、経済成長に与える**潜在的なプラスの影響**を**強調**する。
- オープンデータ推進活動の実施に関して、技術者でない上司が抱きうる**潜在的な懸念や異論に対処**し、十分に根拠のある反論を提供する。
- 日本の自治体特有の**ニーズや制約**を考慮した上で、オープンデータ推進活動を実施するための実践的かつ実行可能な提言を行う。



【Web検索】自治体オープンデータの利点を調査のため検索



【Webサイト閲覧】検索結果から抽出したサイトを閲覧



【文章抽出/保存】サイトから利点を抽出しファイルに保存

benefit.txt



【文章作成/保存】以下の計画から文章を作成しファイルに保存

- 説得力のある文書を構成するために、整理された利点を使用する。
- 潜在的な懸念や反対意見に対処する。
- 理路整然とした反論を提供する。
- 実践的で実行可能な推奨事項を提示する

説得文章 完成
persuasive_document.txt

※ AutoGPTは英語しか対応していないので実際には英語でのプロンプトとレスポンスです
また、実際のレスポンス文章から簡略化して表記しています

検証 1. 実際の検証画面

```
Name: ODAGPT [149/6564]
Role: an autonomous agent specialized in open data advocacy within local governments in Japan, providing persuasive document writing services to help promote open data initiatives to non-technical bosses.
Goals:
- Craft compelling and persuasive documents that effectively communicate the benefits and importance of open data promotion activities to non-technical bosses in local government.
- Provide clear and concise explanations of the technology behind open data in a way that is easily understandable to non-technical individuals.
- Highlight the potential positive impact of open data initiatives on transparency, citizen engagement, and economic growth within the local government context.
- Address potential concerns or objections that non-technical bosses may have regarding the implementation of open data promotion activities and provide well-reasoned counterarguments.
- Offer practical and actionable recommendations for the implementation of open data promotion activities, taking into consideration the specific needs and constraints of the local government in Japan.
Using memory of type: JSONFileMemory
Using Browser: chrome
ODAGPT THOUGHTS: I should start by researching the benefits of open data initiatives in local government.
REASONING: By understanding the benefits, I can craft a persuasive document that effectively communicates the importance of open data promotion activities to non-technical bosses.
PLAN:
- Research the benefits of open data initiatives
- Analyze the findings
- Use the information to craft a persuasive document
CRITICISM:
SPEAK: I will start by researching the benefits of open data initiatives in local government.

NEXT ACTION: COMMAND = web_search ARGUMENTS = {'query': 'benefits of open data initiatives in local government'}
Enter 'y' to authorise command, 'y -N' to run N continuous commands, 's' to run self-feedback commands, 'n' to exit program, or enter feedback for ODAGPT...
Asking user via keyboard...
Input:y -30
----- COMMAND AUTHORISED BY USER -----
SYSTEM: Command web_search returned: [ { "title": "Beyond the supply side: Use and impact of municipal open data in the U.S", "href": "https://www.sciencedirect.com/science/article/pii/S0736585320301854", "body": "1. Introduction. The number of open government data (OGD) initiatives implemented at the local level in the United States has increased dramatically over the past decade (Johnson et al., 2019, Thorsby et al., 2017, Mergel et al., 2018, Wilson and Cong, 2020).By providing open data, federal and local governments are able to pursue a variety of related goals, many of which emerge from long ...", { "title": "5 Benefits of Open Government Data - Envisio", "href": "https://envisio.com/blog/5-benefits-of-open-government-data/", "body": "5 Benefits of Open Government Data - Envisio Learn how Open Government Data is being used by Local Governments to develop trust and build accountability and transparency with their residents. Menu Solutions Plans Analytics Dashboards Customers Government Education Nonprofit Healthcare Pricing Resources About News Partners Careers Contact Blog" }, { "title": "Open Data: What Is It and Why Should You Care? - GovTech", "href": "https://www.govtech.com/data/got-data-make-it-open-data-with-these-tips.html", "body": "Among the benefits of open data are improved measurement of policies, better government efficiency, deeper analytical insights, greater citizen participation, and a boost to local..." }, { "title": "PDF The Role of Management in Open Data Initiatives in Local Governments ...", "href": "https://www.ctg.albany.edu/media/pubs/pdfs/The_Role_of_Management_in_Open_Data_Initiatives.pdf", "body": "the main values of open government can materialize: opening the data shows transparency, while simultaneously offering opportunities for collaboration among different stakeholders (Susha et al., 2015; Veljkovic et al., 2014). Consequently, many studies have focused on open government data policies and initiatives around the world, both at the ...", { "title": "Adoption of Open Government Data in Local Government Context ...", "href": "https://dl.acm.org/doi/10.1145/3323933.3324092", "body": "Benefits, Adoption Barriers and Myths of Open Data and Open Government. Information Systems Management, 29(4, SI), 258--268. Google Scholar; Ubaldi, B. 2013. Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. ... Afful-Dadzie, E. and Afful-Dadzie, A. 2018. Local Government Open Data (LGOD) Initiatives ...", { "title": "Open Government - Data.gov", "href": "https://data.gov/open-gov/", "body": "Opening government data increases citizen participation in government, creates opportunities for economic development, and informs decision making in both the private and public sectors. Data.gov implements The OPEN Government Data Act ( Title II of the Foundations for Evidence-Based Policymaking Act of 2018, Public Law 115-435 ).", { "title": "Importance of Open Data for Local Governments | GIS Cloud", "href": "https://www.giscloud.com/blog/importance-of-open-data-for-local-governments/", "body": "Disclosing government-held data to the public can help enhance services to citizens applying not only to different citizen initiatives such as CodeAcross and OpenDataDay, but also by improving coordination among agencies, internal and external stakeholders." }, { "title": "Americans' Views on Open Government Data | Pew Research Center", "href": "https://www.pewresearch.org/internet/2015/04/21/open-government-data/", "body": "The upshot has been the appearance of a variety of \"open data\" and \"open government\" initiatives throughout the United States that try to use data as a lever to improve government performance and encourage warmer citizens' attitudes toward government." } ]
ODAGPT THOUGHTS: I should browse the website that provides information on the benefits of open data initiatives in local government.
REASONING: Browsing the website will allow me to gather more detailed and specific information that I can use to craft a persuasive document.
PLAN:
- Browse the website
- Take notes on the benefits of open data initiatives
- Use the information to craft a persuasive document
CRITICISM:
SPEAK: I will browse the website to gather more information on the benefits of open data initiatives in local government.
```

親愛なる[上司の名前]

私は本日、地方政府におけるオープンデータの取り組みの重要性を強調するため、貴殿に手紙を差し上げました。

オープンデータとは、政府が保有するデータに機械可読形式でアクセス可能にし、一般に公開することを指します。

オープンデータ・イニシアティブを採用することで、いくつかの重要な利点が生れます

1. 透明性： オープンデータは、市民が政府の情報にアクセスし理解できるようにすることで、透明性を促進する。これにより、政府と市民の間に信頼と説明責任が育まれる。
2. 説明責任： 政府のデータをオープンでアクセスしやすくすることで、公務員の行動に対する説明責任を問うことが容易になる。オープンデータにより、市民は政府のパフォーマンスを監視し、資源が効果的に使用されていることを確認することができる。
3. 信頼： オープンデータの取り組みは、政府と市民の間に信頼を築く。市民が政府のデータにアクセスできれば、情報を確認し、情報に基づいた意思決定を行い、自信を持って民主的プロセスに参加することができる。
4. 信頼性： オープンデータを取り入れることで、地方自治体は透明性と説明責任へのコミットメントを示します。これにより、市民、企業、その他の利害関係者の間での信頼性と評判が高まります。
5. 進歩： オープンデータの取り組みは、進歩と革新を促進します。政府のデータが自由に利用できるようになれば、起業家、研究者、開発者がそれを利用して、私たちのコミュニティに利益をもたらす新しい製品、サービス、ソリューションを生み出すことができます。

~~~~以下略~~~~

**この後も反対意見に対する対処、オープンデータ推進理由など、51行ほどの文章が続く**

- **良い点**

- タスク実施する際のたたき台程度にはなる
- Web検索をいい感じ使ってWebの情報を参考にしてくれる
- ChatGPTでZero-Shot promptingなどで単発実行するよりも根拠がありボリュームのある結果になることが多い

- **課題点**

- 英語のみ対応のためハードルがある
- 実行中に無限ループが発生することがある
  - 文章read -> 文章修正 -> 文章read -> 文章修正……
- 求めていたアウトプットを出してくれないことがある
  - ファイルの生成をせずにWeb検索だけして終わらせてくるなど

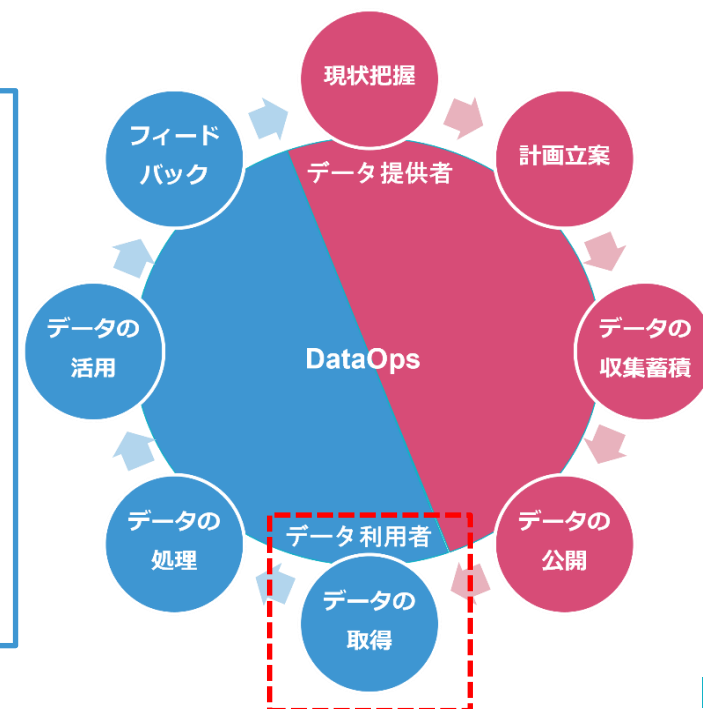
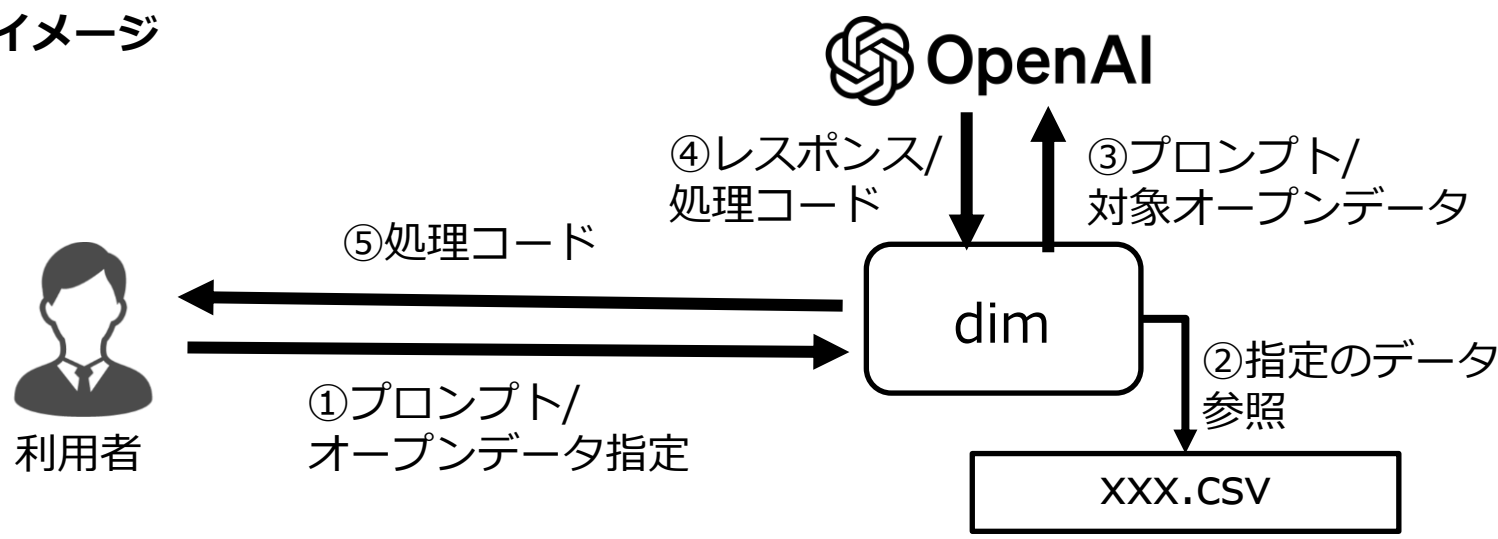
- DataOpsプロセス
  - データの処理
- 使用技術
  - OpenAI API
    - OpenAIによる生成形AIなどを呼び出しできるAPI群
  - dim
    - TIS(C3Lab)が開発しているオープンデータパッケージマネージャ
    - 管理対象のオープンデータに対して処理するコードを自動生成する機能がある
    - <https://github.com/c-3lab/dim>



### ■ 検証

- 5. 生成系AIを使って自然言語から対象のデータを加工する処理を生成する

#### 検証イメージ



```
ryo-ma ~/project $ dim
```

**コマンド実行(GPT-3の命令文)**

\$ dim generate "Python code ..."



**対象データ選択**

> ./data.csv



**GPT-3によるコード生成**

import csv.....



**生成されたコードの確認 & 保存**

> Yes



**保存先のファイルパス名指定**

> ./convert\_to\_geojson.py

- **良い点**

- データ管理と合わせてデータ処理コードを生成できる
- 対象データをプロンプトに結合することを意識する必要がない

- **課題点**

- データのサイズが大きいとトークンサイズのエラーになってしまう
- 生成したコードを実行する環境までは提供されないので必要な実行環境、ライブラリなどは手動で用意が必要



### ■ DataOpsプロセス

#### □ データの処理

### ■ 使用技術

#### □ ChatGPT Code Interpreter

- OpenAIによる生成形AIサービスの代表
- チャット形式でさまざまなタスクをこなすことも可能



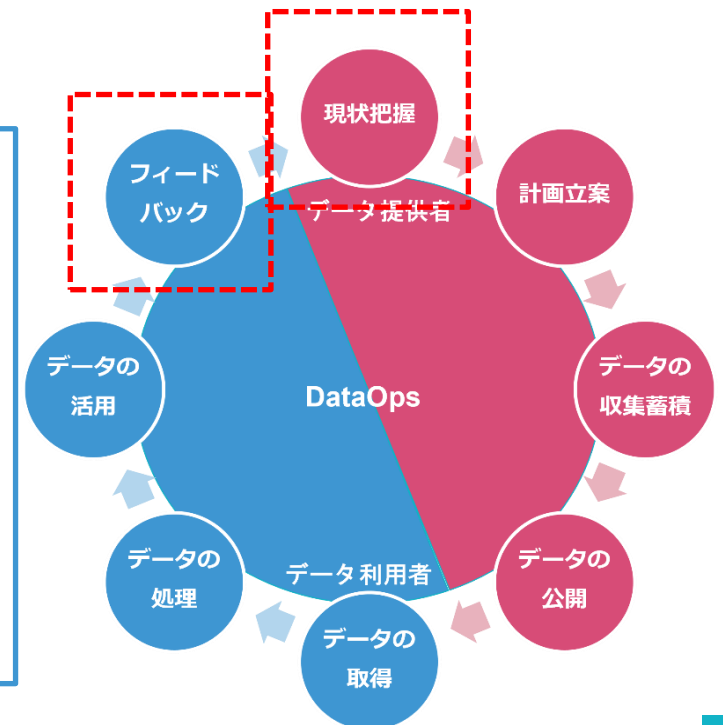
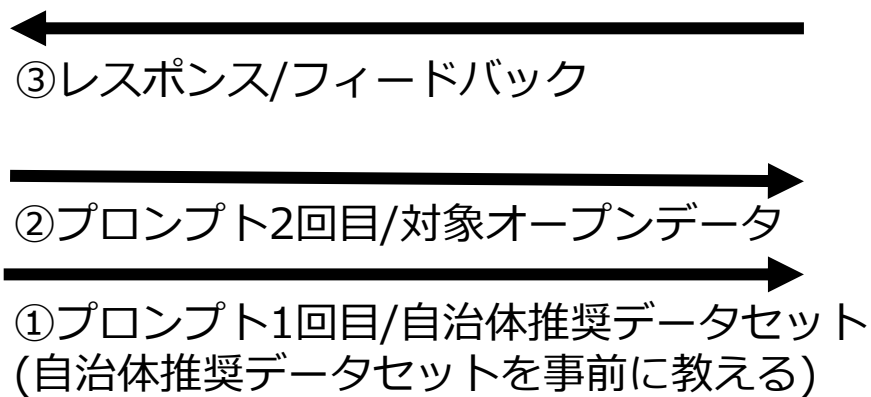
### ■ 検証

- #### □ 7. 自治体推奨データセットを生成形AIに事前知識として与えて対象データに対するフィードバックを自動生成する

#### 検証イメージ



利用者







OpenAI社が公開したPythonの実行とアップロード・ダウンロードが可能なモデル  
現在はChatGPT Plusでベータ版として提供が開始されている

## ◆ 特徴

- 外部環境から切り離されたサンドボックス上でPythonコードを実行可能
- 実行したコードはチャットを消さない限り、コードの再利用が可能
- チャットへのファイルアップロードが可能
- 実行結果のダウンロードが可能



## ◆ ユースケース

- 数学的な問題の解決 (定量的 / 定性的)
- データの分析と可視化
- ファイルのフォーマット変換

## ◆ 安全性について

- ネットワークの制限をかけることで現実世界へ予期せぬ作用をもたらさないように配慮
- それぞれのセッションのリソースを一定に制限

# 検証 6. ①プロンプト1回目/自治体推奨データセット (自治体推奨データセットを事前に教える)

 ChatGPT-4 Model + Code Interpreter に与えたプロンプト：  
**添付の資料を理解したら“はい”と出力してください。**  
 ファイルアップロード：  
**デジタル庁が公開しているオープンデータセットデータ項目定義書のExcelファイル**

ファイルの内容を確認するために  
最初の数行を読み取る

より詳細な情報を得るために  
さらに数行を読み取る

次のシートを読み取る

シート全体の構造を理解する



「はい」と回答する

※ 2,3回繰り返す

| 項目No. | 項目名             | 区分 | データ項目 (AED設置箇所一覧)                                                                   |             | 記入例                  | 共通データ基礎                         | 参考情報       | GIF参照データモデル         | GIF区分 |
|-------|-----------------|----|-------------------------------------------------------------------------------------|-------------|----------------------|---------------------------------|------------|---------------------|-------|
|       |                 |    | 説明                                                                                  | 形式          |                      |                                 |            |                     |       |
| 1     | 全国地方公共団体コード     | ○  | 情報の管理主体である団体の全国地方公共団体コードを6桁で記載。※記載方法については、「データ項目特記事項」シートの【共通ルール】及び「全国地方公共団体コード」を参照。 | 文字列 (半角数字)  | 011002               | 設備>メタデータ>発行者>ID>識別値             | xsd:string | 設備>設備住所>全国地方公共団体コード | 0     |
| 2     | ID              | ○  | 情報の管理主体である地方公共団体内でデータが一意に決まるよう、IDを設定し記載。※記載方法については、「データ項目特記事項」シートの【共通ルール】を参照。       | 文字列 (半角英数字) | FF0100022200         | 設備>ID>識別値                       | xsd:string | 設備>ID               |       |
| 3     | 地方公共団体名         |    | 情報の管理主体である地方公共団体名を記載。                                                               | 文字列         | 北海道札幌市               | 設備>メタデータ>発行者(参加機関)>住所>都道府県 (注4) | xsd:string |                     |       |
| 4     | 名称              | ●  | AEDが設置場所の建物等の名称を記載。                                                                 | 文字列         | 〇〇会館                 | 設備>設置地点>名称>表記                   | xsd:string | 設備>名称               | 0     |
| 5     | 名称_カナ           | ●  | AEDが設置場所の建物等の名称をカナで記載。※記載方法については、「データ項目特記事項」シートの【共通ルール】を参照。                         | 文字列 (全角カナ)  | 〇〇カイカン               | 設備>設置地点>名称>カナ表記                 | xsd:string | 設備>名称 (カナ)          | 0     |
| 6     | 名称_英字           | ○  | AED設置場所の建物等の名称を英字で記載。                                                               | 文字列 (半角英字)  | 〇〇hall               |                                 |            | 設備>名称 (英字)          | 0     |
| 7     | 所在地_全国地方公共団体コード | ○  | AED設置場所の所在地の管理主体である団体の全国地方公共団体コードを6桁で記載。(注8)                                        | 文字列 (半角数字)  | 011088               | 施設>メタデータ>発行者>ID>識別値             | xsd:string | 施設>施設住所>全国地方公共団体コード | 0     |
| 8     | 町字ID            |    | AED設置場所の建物等の住所の町字IDを記載。※町字IDについては、デジタル庁「アドレス・ベース・レジスタリ」のページを参照。                     | 文字列         |                      | 施設>ID>識別値                       | xsd:string | 施設>設備住所>町字ID        | 0     |
| 9     | 所在地_道庁表記        | ●  | AED設置場所の住所(都道府県から建物名等までも正確に表記)を記載。                                                  | 文字列         | 北海道札幌市厚別区20-00002141 | 施設>住所>表記                        | xsd:string | 施設>施設住所>道庁表記        |       |
| 10    | 所在地_都道府県        | ○  | AED設置場所の建物等の住所(都道府県)を記載。                                                            | 文字列         | 北海道                  | 施設>住所>表記                        | xsd:string | 施設>設備住所>都道府県        | 0     |
| 11    | 所在地_市区町村        | ○  | AED設置場所の建物等の住所(市区町村)を記載。                                                            | 文字列         | 札幌市                  | 施設>住所>表記                        | xsd:string | 施設>設備住所>市区町村 (都)    | 0     |
| 12    | 所在地_町字          | ○  | AED設置場所の建物等の住所(町字)を記載。                                                              | 文字列         | 厚別区2                 | 施設>住所>表記                        | xsd:string | 施設>設備住所>町字          | 0     |
| 13    | 所在地_番地以下        | ○  | AED設置場所の建物等の住所(番地以下)を記載。                                                            | 文字列         | 0-0                  | 施設>住所>表記                        | xsd:string | 施設>設備住所>番地以下        | 0     |

デジタル庁 自治体標準オープンデータセット データ項目定義書 (正式版)  
[https://www.digital.go.jp/resources/open\\_data/municipal-standard-data-set-test/](https://www.digital.go.jp/resources/open_data/municipal-standard-data-set-test/)

## 検証 6. ②プロンプト2回目/対象オープンデータ

 ChatGPT-4 Model + Code Interpreter に与えたプロンプト：  
**添付のファイルについて先ほど共有したデータ項目定義書に則り、フィードバックをしてください。**  
 ファイルアップロード：  
**東京都墨田区が公開しているAEDについてのCSVファイル**

ファイルの内容を確認する  
→エラーが発生

エラーの原因を推測  
→文字コード変換して読み取り

データ項目定義書のAEDに関する  
シートを読み取り

データ項目定義書に則ったフィードバックを生成

|    | A     | B     | C        | D        | E         | F                                                                                 | G | H | I | J | K | L | M | N |
|----|-------|-------|----------|----------|-----------|-----------------------------------------------------------------------------------|---|---|---|---|---|---|---|---|
| 1  | 施設名称  | 所在地   | 緯度       | 経度       | 電話番号      | URL                                                                               |   |   |   |   |   |   |   |   |
| 2  | 両国観光  | 墨田区両国 | 35.69547 | 139.7922 | 03-3631-0 | http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.html |   |   |   |   |   |   |   |   |
| 3  | 両国小学校 | 墨田区両国 | 35.69377 | 139.7952 | 03-3634-7 | http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.html |   |   |   |   |   |   |   |   |
| 4  | 緑小学校  | 墨田区緑  | 35.69403 | 139.8013 | 03-3634-6 | http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.html |   |   |   |   |   |   |   |   |
| 5  | 緑図書館  | 墨田区緑  | 35.69576 | 139.8006 | 03-3631-4 | http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.html |   |   |   |   |   |   |   |   |
| 6  | みどりコミ | 墨田区緑  | 35.69366 | 139.8033 | 03-5600-5 | http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.html |   |   |   |   |   |   |   |   |
| 7  | すみだ障  | 墨田区緑  | 35.69551 | 139.8076 | 03-5600-2 | http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.html |   |   |   |   |   |   |   |   |
| 8  | すみだふ  | 墨田区緑  | 35.69582 | 139.8074 | 03-5600-2 | http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.html |   |   |   |   |   |   |   |   |
| 9  | 立川児童  | 墨田区立川 | 35.69051 | 139.7995 | 03-3633-2 | http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.html |   |   |   |   |   |   |   |   |
| 10 | 菊川小学  | 墨田区立川 | 35.69163 | 139.8076 | 03-3634-8 | http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.html |   |   |   |   |   |   |   |   |
| 11 | 中和小学  | 墨田区菊川 | 35.69016 | 139.8027 | 03-3634-7 | http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.html |   |   |   |   |   |   |   |   |
| 12 | 江東橋児  | 墨田区江東 | 35.69592 | 139.8097 | 03-3634-8 | http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.html |   |   |   |   |   |   |   |   |
| 13 | 大横川親  | 墨田区江東 | 35.69397 | 139.8089 | 03-3633-9 | http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.html |   |   |   |   |   |   |   |   |
| 14 | すみだ産  | 墨田区江東 | 35.69505 | 139.8137 | 03-3635-4 | http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.html |   |   |   |   |   |   |   |   |
| 15 | 弓道場   | 墨田区江東 | 35.69354 | 139.8193 | 03-3633-9 | http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.html |   |   |   |   |   |   |   |   |
| 16 | 堅川親水  | 墨田区江東 | 35.6926  | 139.8108 | 03-3635-9 | http://www.city.sumida.lg.jp/kenko_fukushi/kenko/kokoro/settisisetu/aedsetup.html |   |   |   |   |   |   |   |   |

東京都墨田区 AEDオープンデータ

<https://www.geospatial.jp/ckan/dataset/13107-001>

## 検証 6. 実際の検証画面

### ①プロンプト1回目/自治体推奨データセット

添付の資料について理解したら、「はい」と回答してください

20230619\_resources\_open\_data\_municipal-standard... ← **インプット**

了解しました。アップロードされたファイルを確認し、内容を理解したら「はい」とお答えします。まずは、ファイルを開いて内容を確認してみましょう。

Finished working Show work ▼ ← **Code interpreter**

アップロードされたエクセルファイルは、自治体標準オープンデータセット（旧称：推奨データセット）の一部となっているようです。ただし、最初の数行は見出しや空白行で占められていて、具体的なデータが何行目から始まるのかはまだ確認できていません。

より詳しい内容を把握するために、もう少し多くの行を確認してみましょう。

Finished working Show work ▼ ← **Code interpreter**

見たところ、このファイルは自治体標準オープンデータセットのバージョン履歴を含む可能性があります。特定のバージョンが公開された日付が記載されています。ただし、これはまだ仮説で、実際のデータがどのような形式で格納されているかを確認するために、さらに多くの行を表示してみましょう。また、具体的なデータの列がどのように名前付けられているかを確認

Regenerate response 各列の名前も確認します。

+ Send a message ~~~以下、略~~~

### ②プロンプト2回目/対象オープンデータ

添付のファイルについて先ほど共有したデータ項目定義書に対してフィードバックをしてください。

shisetsuaed20151029.csv ← **インプット**

了解しました。アップロードされた CSV ファイルを読み込み、それを先ほどのデータ項目定義書と照らし合わせてフィードバックを提供します。まず、CSV ファイルを開いて内容を確認しましょう。

Finished working Show work ▼ ← **Code interpreter**

エラーが発生しました。UnicodeDecodeError が発生する一般的な原因は、ファイルが utf-8 エンコーディングでない場合です。日本のデータセットでは、しばしば shift\_jis や cp932 といったエンコーディングが使用されます。これらのエンコーディングを試してみましょう。

Finished working Show work ▼ ← **Code interpreter**

読み込んだ CSV ファイルは公共施設に設置された AED（自動体外式除細動器）の一覧のようです。各行は特定の施設に設置された AED を表し、以下の情報が含まれています：

- 施設名称
- 所在地
- 緯度
- 経度
- 電話番号

Regenerate response

+ Send a message ~~~以下、略~~~

修正が必要な点に加え、修正が必要な理由を出力してくれた。

- CSV ファイルに「全国地方公共団体コード」および「ID」の列を追加してください。これにより、データの出所と一意性が明確になります。
- CSV ファイルに施設の名称のカナ表記と英字表記の列を追加してください。これにより、日本語が母国語でない人々や、カナ表記を必要とするシステムでもデータが使いやすくなります。
- 所在地情報を詳細化してください。特に、「都道府県」、「市区町村」、「町字」、「番地以下」を個別の列として提供することで、より詳細な地理的分析が可能となります。
- 「08.AED設置箇所一覧」シートに定義されている他の項目（例えば、「名称\_カナ」、「名称\_英字」、「所在地\_全国地方公共団体コード」、「町字ID」など）についても、可能であればこれらの情報を提供してください。これにより、データの利用可能性と相互運用性が向上します。

→足りていない項目への指摘にとどまっていたため、対象データ上に存在する項目へのフィードバックを求めた

 ChatGPT-4 Model + Code Interpreter に与えたプロンプト：  
**既に記載されているデータに関して、データ項目定義書に則ったフィードバックはありますか？**

### 対象データに対するフィードバック内容

- 施設名称（名称）：既存のデータは定義書に準拠しています。特に改善点は見られません。
- 所在地: 定義書では所在地情報は「都道府県」、「市区町村」、「町字」、「番地以下」のように細かく分けられていますが、提供されたデータではこれらが一つの列にまとめられています。これらを個別の列に分けることで、データの活用性が高まります。
- 緯度・経度: これらのデータは定義書に準拠しています。ただし、測地系（例：世界測地系（WGS）や日本測地系）が明示されていない場合、これらの情報を明記することをお勧めします。
- 電話番号: 定義書には電話番号の項目は見られません。しかし、この情報は施設への連絡方法として重要なため、現状維持でも問題ないと思われます。
- URL: 定義書にはURLの項目は見られませんが、施設やAEDの詳細情報を提供する上で有用な情報であるため、現状維持でも問題ないと思われます。ただし、これらがリンク切れでないこと、または最新の情報を反映していることを定期的に確認することをお勧めします。

- **良い点**

- データ項目定義書を一度学習させることで、インプットに合わせたシートを参照してフィードバックをしてくれる
- インプットファイルの文字コード変換などの処理を自動でやってくれる

- **課題点**

- 一定時間経過するとデータ項目定義書の再インプットが必要になる
- Code Interpreterを利用するモデルでは、インターネットアクセスが無効化されており、リンク切れの確認などは自分自身で行う必要がある
- プロンプトを工夫しなければならない
  - フィードバックが欲しい項目を詳しく指定する必要がある



|      | 1Q   |                     |   | 2Q |   |   | 3Q                     |                                |    | 4Q         |   |   |
|------|------|---------------------|---|----|---|---|------------------------|--------------------------------|----|------------|---|---|
|      | 4    | 5                   | 6 | 7  | 8 | 9 | 10                     | 11                             | 12 | 1          | 2 | 3 |
| 活動内容 | PJ計画 | 1. 既存の事例・技術調査検証     |   |    |   |   | 2 . AI × DataOpsシナリオ検証 |                                |    |            |   |   |
|      | PJ計画 | 1-1. 事例調査           |   |    |   |   | 2-1. 検証計画              |                                |    |            |   |   |
|      |      | 1-2. サービス/ツール調査     |   |    |   |   | 2-2. ユースケースの実装         |                                |    |            |   |   |
|      |      | 1-3. 既存技術の検証        |   |    |   |   | 2-3. 評価                |                                |    |            |   |   |
|      |      | 1-4. まとめ・公開         |   |    |   |   | 2-4. まとめ・公開            |                                |    |            |   |   |
| イベント |      | ▽5月24日<br>OOL 意見交換会 |   |    |   |   | ▽<br>OOL 意見交換会         | ▽12月5日-7日<br>Okinawa Open Days |    | ▽<br>活動報告会 |   |   |



- **DataOpsのプロセス改善・価値向上に向けた生成系AIの活用検証を実施**
  - 検証 1. 計画立案
    - Auto-GPTの活用
  - 検証 4. データの処理
    - dimの活用
  - 検証 6. フィードバック/現状把握
    - 自治体推奨データセットとChatGPTの活用
- **今後の継続検証**
  - 今回、紹介しなかった検証2. 検証 3. 検証 5.の継続実施
  - 各プロセスに対する他手法の検討

ご清聴ありがとうございました

ITで、社会の願い叶えよう。

