

MATLAB EXPO

30, May, 2024 | Japan

知識を持って挑む

生成AIの活用と自然言語処理による業務効率化

Misa Taguchi / MathWorks Japan



(She/Her)

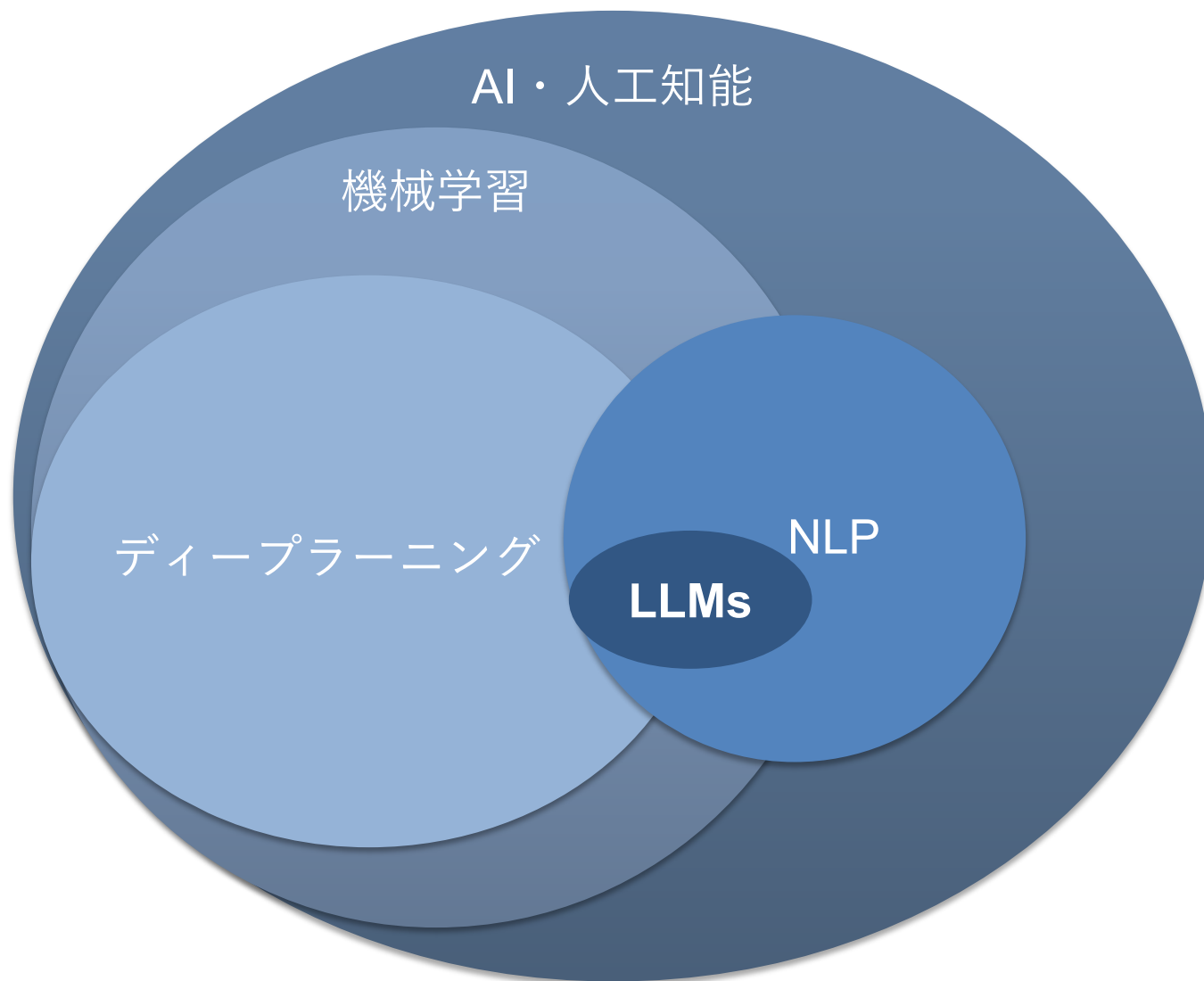


アジェンダ

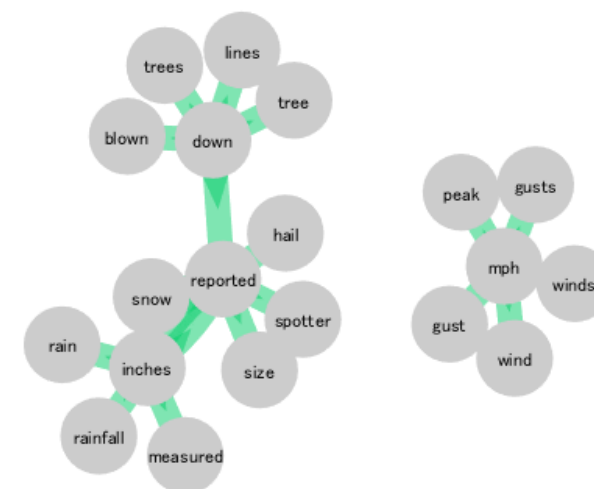
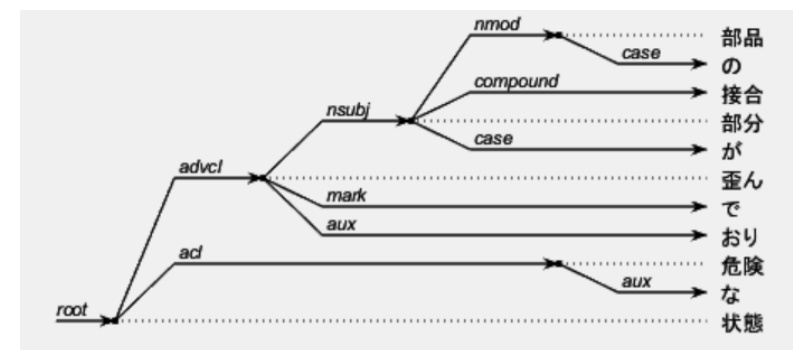
- 生成AI・自然言語処理とは
- 産業界における自然言語処理活用例
- LLMsの活用例 – チャットボットシステム
- MATLABでアクセス可能なLLMs

生成AI・自然言語処理とは

自然言語処理（NLP）・大規模言語モデル（LLMs）とは

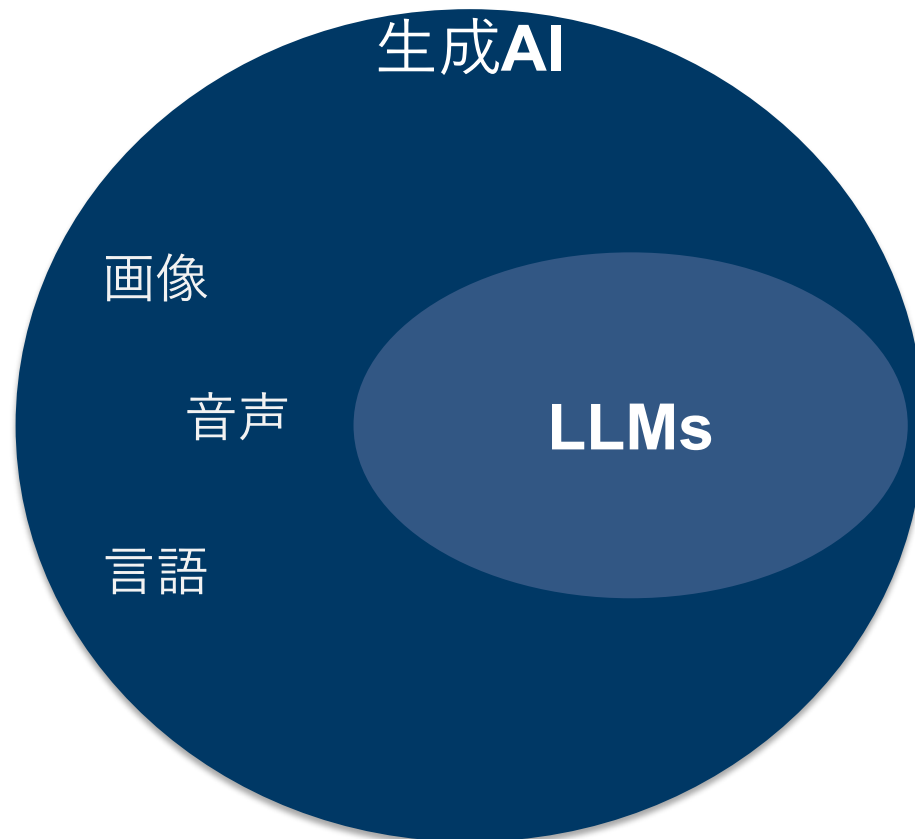


- NLP = 言語データで世の中の問題を解決



生成AIとは

- 大量の既存データを学習し、そのパターンや特徴を基に新しいコンテンツ（画像、音声、言語など）を生成する人工知能技術

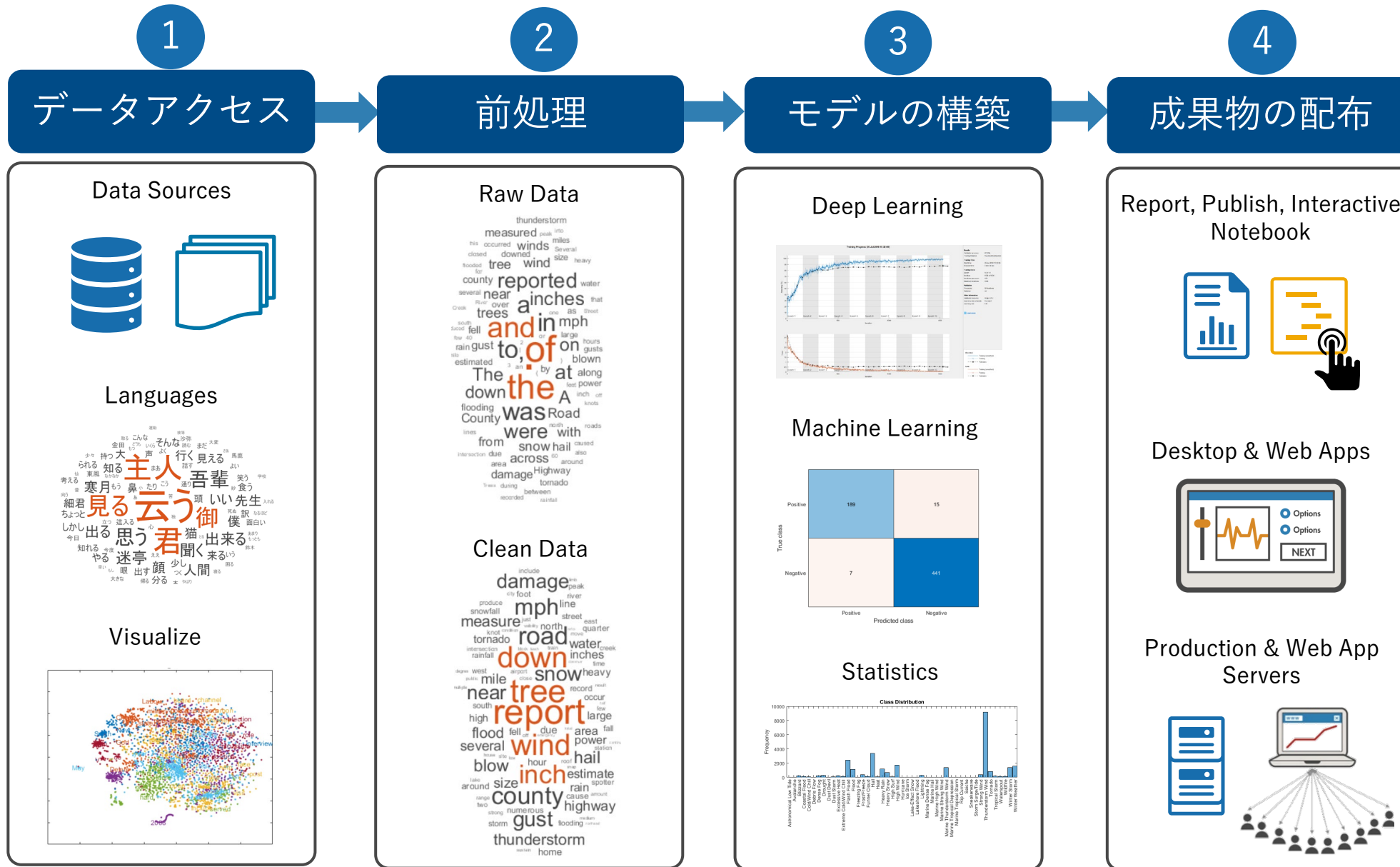


敵対的生成ネットワーク
(GAN) にて作成した画像

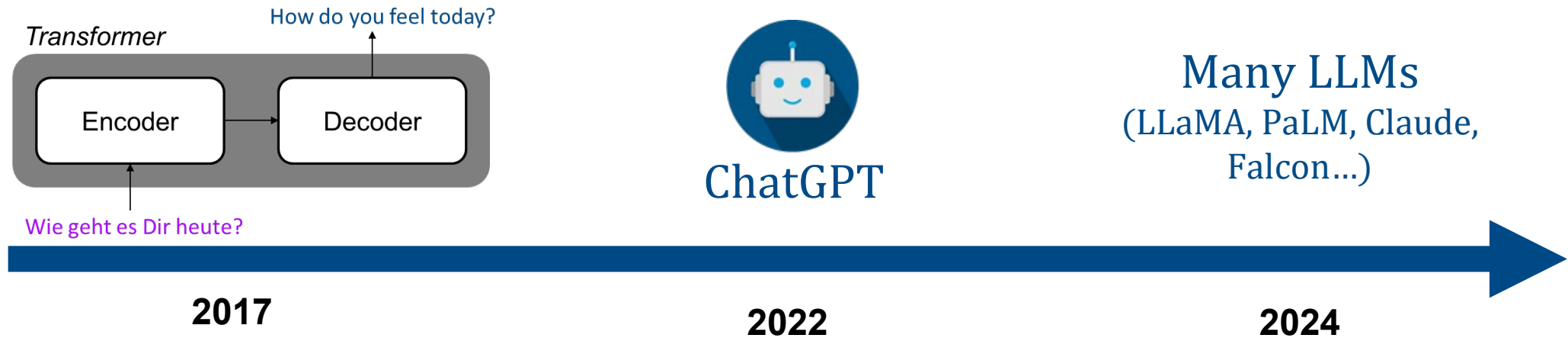


例題

自然言語処理における基本ワークフロー



LLMsの発展と利用のメリット



- 要約作成など自然言語処理タスク精度が著しく向上
- 前処理不要でモデルが利用可能に

デモブースにて
体験いただけます！

LLMs活用のための3つのアプローチ方法

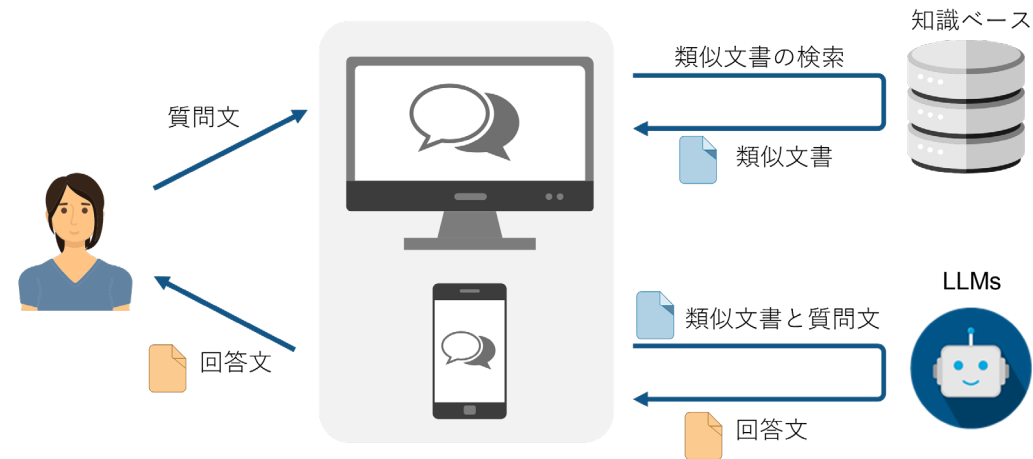
①



Prompt Engineering

精度の高いアウトプットが得られるように質問文を構築する

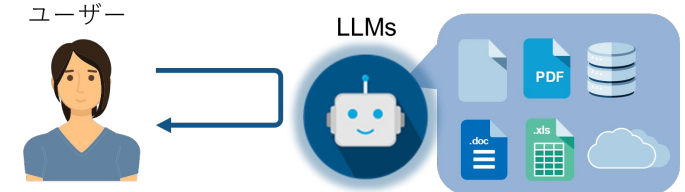
②



Retrieval-Augmented Generation (RAG)

回答が含まれていそうなテキストを質問文と一緒にLLMに渡して、LLMが学習していない内容でも回答を得られるようにする

③



Fine-tuning

ドメイン固有のテキストデータを使ってLLMの重みを微調整（再学習）する

小まとめ

- 自然言語処理はAIの一部
- LLMsの台頭により精度や使い勝手が向上した
- LLMs活用のために3つのアプローチが存在する
 - プロンプトエンジニアリング
 - RAG
 - ファインチューニング

産業界における自然言語処理活用例

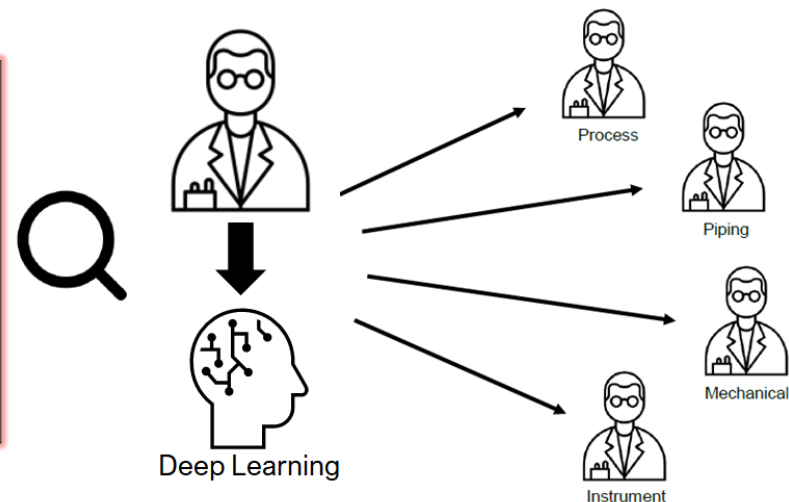
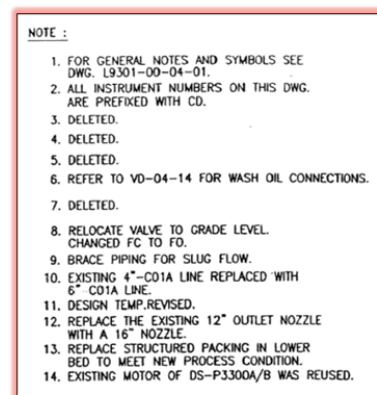
サムスンエンジニアリング、P&IDノート分類のための テキスト分析深層学習モデルを開発

課題

P&IDには、エンジニアがプラント設計時に実施しなければならない特別な要件が記載されており、専任のエンジニアはそのノートを適切な分野に分類する必要があります。このようなノートの確認や分類には数週間を要し、既存のルールベースのプログラムではすべてのケースをカバーすることはできません。

ソリューション

サムスンエンジニアリングは、人のレベルと同等の精度で、一貫性を保証しながらP&IDノートをレビューできるよう、深層学習ベースのLSTMテキスト分類モデルを開発しました。



MATLABとSimulinkを使用する利点

- テキストデータ前処理のための高水準API
- 実験マネージャーによる効率的なハイパーパラメータチューニング
- 実践で使えるドキュメント例

“ 深層学習はエンジニアの作業を高速化するのに役立ち、
MATLABはそれをさらに高速化することができます ”

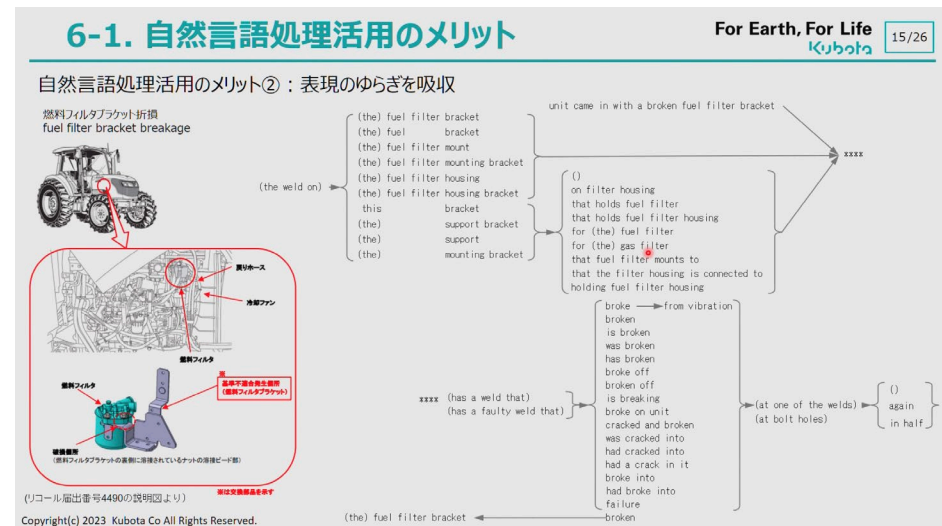
Geun-yong Choi, サムスンエンジニアリング

クボタ、故障対応の申請内容を自然言語処理技術で分析し、故障メカニズムを素早く推定

自然言語処理技術とMATLABを活用して故障対応の申請内容から類似度スコアを算出し、素早い故障メカニズムの推定や効率のよい案件の判別を実現しました。

主な成果/利点:

- MATLABを使ったツールを開発したことで、ワランティ読解時間を1/3以下に短縮したケースも
- ツール改修の際もコードを書き換えることでフレキシブルな改修が可能に。アプリ化の際、App Designerを用いて画面デザインも自分たちで作成できるように
- 既存のPythonモジュールをそのまま利用でき、MATLABによる開発環境へ容易に移行が可能に。今後は積極的な活用を目指す



自然言語処理活用のメリットは、意味を解釈し一発で分類できることや、表現のゆらぎを吸収できることにある

大量データを処理したり、試行錯誤を繰り返して評価アルゴリズムを構築するツールとしてMATLABに着目。マトリクスへの扱いやすさ、処理をコンパクトに記述できる点、ユーザーインタフェースの柔軟性、インタラクティブな操作性といったMATLABの特長を行かして取り組みを進めました。

LLMsの活用例 – Q&Aタスク

チャットボット作成のための3つのアプローチと手法毎の長所短所

手法	データセット	事前学習	インターネット接続	ハルシネーション
1. Q&Aのペアを事前準備	要	<u>不要</u>	<u>不要</u>	<u>なし</u>
2. BERT利用 (抽出型)	要	要	<u>不要</u>	<u>なし</u>
3. ChatGPT利用 (生成型)	<u>不要</u>	<u>不要</u>	要	あり

抽出型 (Extractive) と生成型 (Generative)

- 要約作成タスクを例とした抽出型AIと生成型AIの比較

Text Analytics Toolbox R2024a

テキストデータの解析とモデル化

リリースノート

PDF版ドキュメンテーション

Text Analytics Toolbox™ は、テキストデータの前処理、解析、モデル化のためのアルゴリズムと可視化を提供します。ツールボックスで作成されたモデルは、センチメント分析、予知保全、トピックモデリングなどの用途で使用できます。

Text Analytics Toolbox には、装置のログ、ニュースフィード、アンケート、オペレーターレポート、ソーシャルメディアなどのソースから得た生テキストを処理するためのツールが含まれます。一般的なファイル形式からテキストを抽出し、生テキストを前処理し、個々の単語を抽出し、テキストを数値表現に変換し、統計モデルを構築することができます。

LSA、LDA、単語埋め込みなどの機械学習手法を使用して、クラスターを検出し、高次元のテキストデータセットから特徴を作成することができます。Text Analytics Toolbox で作成された特徴を他のデータソースの特徴と組み合わせて、テキスト、数値、およびその他の種類のデータを利用する機械学習モデルを構築することができます。

抽出型



元の文から抽出・不自然な要約になるリスクも

Text Analytics Toolbox™ は、テキストデータの前処理、解析、モデル化のためのアルゴリズムと可視化を提供します。

生成型

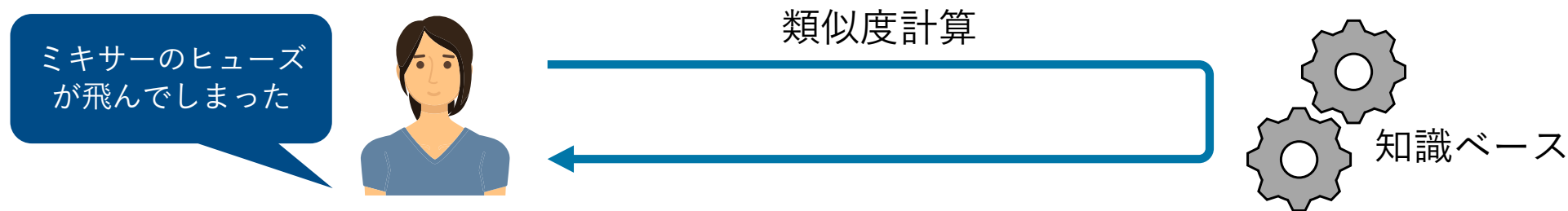


AIが文章を作文・ハルシネーションのリスクも

Text Analytics Toolbox™ は、テキストデータの前処理、解析、モデリングを提供し、センチメント分析やトピックモデリングなどの用途に活用可能なツールです。

1. 事前にQ&Aのペアを用意しておく方法

- 事前に用意したQ&Aのペアの中から一番類似しているペアを検索する



不具合の詳細 (Question)	解決方法 (Answer)
'スキャナーのスプールにアイテムが詰まることがある。'	'マシンの再調整'
'アセンブラのピストンからガタガタと大きな音がします。'	'マシンの再調整'
'工場起動時に電源が切れる。'	'フル交換'
'アセンブラのコンデンサが飛ぶ。'	'部品交換'
'ミキサーでヒューズが切れる。'	'ウォッチリストに追加'
'コンストラクション・エージェントのパイプが破裂して冷却水を噴射して...'	'部品交換'
'ミキサーでヒューズが飛んだ。'	'部品交換'

2. BERT*を用いた方法（抽出型）

*Bidirectional Encoder Representations from Transformers

- [質問]と[回答が含まれる文]をセットで送り、実際の回答となる文節位置情報を得る

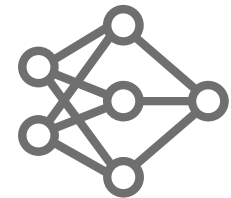


[質問] When is her birthday?

[回答が含まれる文] July 24 in 1999,
Yoko was born at Saiseikai hospital.



BERT



[回答が含まれる文] **July 24 in 1999,**
Yoko was born at Saiseikai hospital.

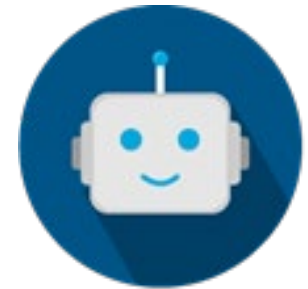
事前学習は要するが、オフラインで文脈を考慮した回答を得られる

3. ChatGPTを用いた方法（生成型）

- ChatGPTが事前学習していない内容はプロンプトに含めて質問することも可能



- 質問文
- 質問文 + 回答が含まれる文



Web API接続が必要だが、事前学習不要で、新しい質問にも柔軟に対応可能

小まとめ

- 抽出型と生成型のモデルが存在する
- 各モデルの長所・短所を理解した上での方針策定を推奨

手法	データセット	事前学習	インターネット接続	ハルシネーション
1. Q&Aのペアを事前準備	要	<u>不要</u>	<u>不要</u>	<u>なし</u>
2. BERT利用（抽出型）	要	要	<u>不要</u>	<u>なし</u>
3. ChatGPT利用（生成型）	<u>不要</u>	<u>不要</u>	要	あり

MATLABでアクセス可能なLLMs

MATLABで利用可能なLLMs

BERT

Sentence
BERT

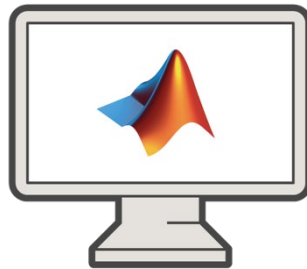
R2024a

FinBERT

GPT-2

Tohoku
BERT

日本語モデル

[トランスフォーマーモデルGitHub](#)

ChatGPT

*OpenAIと接続

[Call OpenAI from MATLAB GitHub](#)

MATLABだから叶う ローコードLLMs活用例

- GPT-2によるオフライン環境での要約作成

```
mdl = gpt2;  
inputText = [要約したい文書];  
generateSummary(mdl, inputText)
```

- BERTによるローコードなファインチューニング

```
mdl = bertDocumentClassifier(ClassNames=クラス名);  
mdl = trainBERTDocumentClassifier(テキスト, ラベル  
, mdl, options);
```

AI Chat Playground - MATLAB コードに最適化されたテキスト生成AI

2023年11月公開

MathWorks 製品 ソリューション アカデミア サポート コミュニティ イベント

MATLAB Answers File Exchange Cody AI Chat Playground Discussions Contests Blogs More

New

Copy Download

AI

Hello! Whether you're checking out MATLAB for the first time or have been using it for years, I'm here to answer your questions and provide coding tips.

Please keep in mind that AI sometimes writes code and text that seems accurate, but isn't. AI does not yet have knowledge of features delivered after June 2021 and only limited knowledge of Simulink and specialized toolboxes. This is a space for experimentation. Try it, verify any resulting code, and kindly give or on the results to help improve the responses.

Read "peppers.png" and display as grayscale

質問を入力 (日英いずれも可)

geoscatteer to visualize

Shuffle

Enter prompt here

Generated code and text might be inaccurate. Verify before use. @About @Dislikes

CODE RUN

画像の読み込み、可視化、グレースケール化を

小まとめ

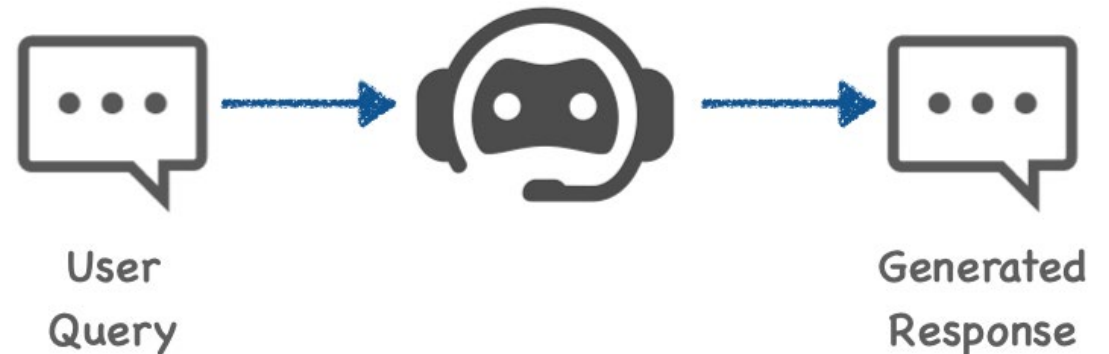
- オフラインのLLMsにアクセスしセキュアな解析が可能
- ローコードでLLMsを効率良く活用
- AI Chat Playgroundによるコーディング作業の効率化

Key Takeaways

- 生成AI・自然言語処理とは
- 産業界における自然言語処理活用例
- LLMsの活用例 – チャットボットシステム
- MATLABでアクセス可能なLLMs

- ブース出展中！ぜひお立ち寄りください

試してみよう！
MATLABで大規模言語モデル



MATLAB EXPO

知識を持って挑む

生成AIの活用と自然言語処理による業務効率化

Thank you



© 2024 The MathWorks, Inc. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See [mathworks.com/trademarks](https://www.mathworks.com/trademarks) for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.

