

at-10. 自然言語処理の基礎

(ディープラーニングのシステムとプログラミング)
(全12回)

<https://www.kkaneko.jp/ai/at/index.html>

金子邦彦



- 
- ① 深層学習（ディープラーニング）の活用により高度な言語処理が可能に
 - ② 日本語を含む多言語の処理に Python が有効である
 - ③ 情報検索、対話、翻訳、要約に加え、OCR や音声合成など応用範囲が広い
 - ④ 画像からテキスト抽出が可能で、多様なデータを扱える

自然言語処理

自然言語処理は、人間が普段使う言語（日本語、英語など）をコンピュータが理解したり、生成する技術

自然言語処理の用途

用途1: 言語の効果的な処理

例: 翻訳システム（Google翻訳、DeepL翻訳など）は、異なる言語間で翻訳

用途2: 人間との自然な対話の実現

例: スマートフォンのアシスタント（Siriやアレクサ）は、ユーザーの質問に答えるなど、人間との自然な対話を行う

日常生活で使われ、便利なサービスや製品が登場

「明日の天気は」

KU

You

明日の天気は？



ChatGPT

明日の天気について、以下の情報が得られました：

- **全国的な予報:** 明日は前線が通過し、大気の状態が不安定になる見込みです。北海道や東北では日本海側を中心に雨や雪が予想され、一部地域では吹雪く可能性もあります。北陸地方では断続的に雨や雷雨が発生し、突風やひょうにも注意が必要です。関東から九州、沖縄地方では昼頃まで雨や雷雨の可能性がありますが、日中は広く晴れる見込みです。最高気温は平年並みかやや高めですが、午後からは気温が下がることが予想されます。西日本と沖縄では黄砂に注意が必要です [1]。

AI の回答

この予報は全国的な概観であり、地域によって異なる可能性があります。詳しい地域別の天気予報については、気象情報サイトやニュースアプリで確認してください。

人間の言葉を使用しての AI とのコミュニケーション

自然言語処理のさまざまな応用

- **情報検索**：インターネットから情報を検索
例：検索エンジンでの、キーワードや文章による検索
- **AIとの対話、AIへの指示**：AIが対話や指示に応じる
例：AIアシスタント（Siriやアレクサ）
- **プログラミング支援**
- **人間の指示による文書の作成や推敲**
- **翻訳**：異なる言語間で翻訳
例：動画の自動翻訳、字幕付け
- **要約**：長文からポイントを抜き出す

検索エンジンでの、キーワードや文章による検索

例：「サステナビリティ」で検索

Google サステナビリティ

画像 ニュース SDGs 違い 簡単に 取り組み 企業 経営 動画 取り組み事例

すべてのフィルタ ツール セーフ

約 131,000,000 件 (0.30 秒)

サステナビリティ (sustainability) とは、環境や経済等に配慮した活動を行うことで、社会全体を長期的に持続させていこうという考えです。「Sustain (維持する、持続する)」と「Ability (～する能力)」を組み合わせた造語で、日本語では「持続可能性」と呼ばれます。 2023/11/14

三井住友銀行
<https://www.smbc.co.jp/planning/about-sustainability> :
サステナビリティとは？企業が取り組むメリットや指標

サステナビリティ 3つの柱

- 環境保護 (Environmental Protection)
- 社会開発 (Social Development)
- 経済発展 (Economic Development)

持続可能性

持続可能性は、将来にわたって現在の社会の機能を継続していくことができるシステムやプロセスのこと。一般的にはそういった仕組みを指すが、環境学的には生物学的システムがその多様性と生産性を期限なく継続できる能力のことを指し、さらに組織原理としては、持続可能な発展を意味する。 [ウィキペディア](#)

他の人はこちらも検索 他 5 件以上を表示

AI との対話



あなた

週末に広島県尾道市に行きますが、楽しいアクティビティを1つ簡単に紹介してください

ChatGPT

尾道市での楽しいアクティビティとしては、尾道水道を渡る「尾道フェリークルーズ」がおすすめです。尾道水道に架かる東西2つの橋をくぐりながら、尾道の風景を船上から楽しむことができます。船から見る尾道の街並みや美しい風景は絶景で、のんびりとした時間を過ごせます。船に乗りながら尾道の魅力を感じることができるので、訪れた際にはぜひ試してみてください。

AI の回答



あなた

料金はどこで調べたらいいですか？

ChatGPT

料金を調べるには、以下の方法があります。

1. インターネットで検索する: 商品やサービス名をキーワードにして、比較サイトやオンラインショップなどで料金を検索することができます。
2. 公式ウェブサイトをチェックする: 商品やサービスを提供している企業や店舗の公式ウェブサイトを訪れて、料金やプランなどの情報を探することができます。
3. お問い合わせする: 商品やサービスを提供している企業や店舗に直接お問い合わせをすることで、料金などの詳細を確認できます。
4. カタログやチラシをチェックする: 店舗や企業が提供しているカタログやチラシには、料金やプランなどの情報が掲載されていることがあります。
5. 友人や知人に相談する: 同じ商品やサービスを利用している友人や知人に料金について相談することで、参考になる情報を得ることができます。

以上の方法を組み合わせて、自分が調べたい料金情報を見つけることができます。

AI の回答

AI への指示



AIが対話や指示に応じる

- ・ 音声による家庭内のさまざまな装置の操作
- ・ 音声コマンドによるスマホや、スマートスピーカーの操作

AI への指示 (文章の指示による画像の合成)

I like chocolate

元の文章

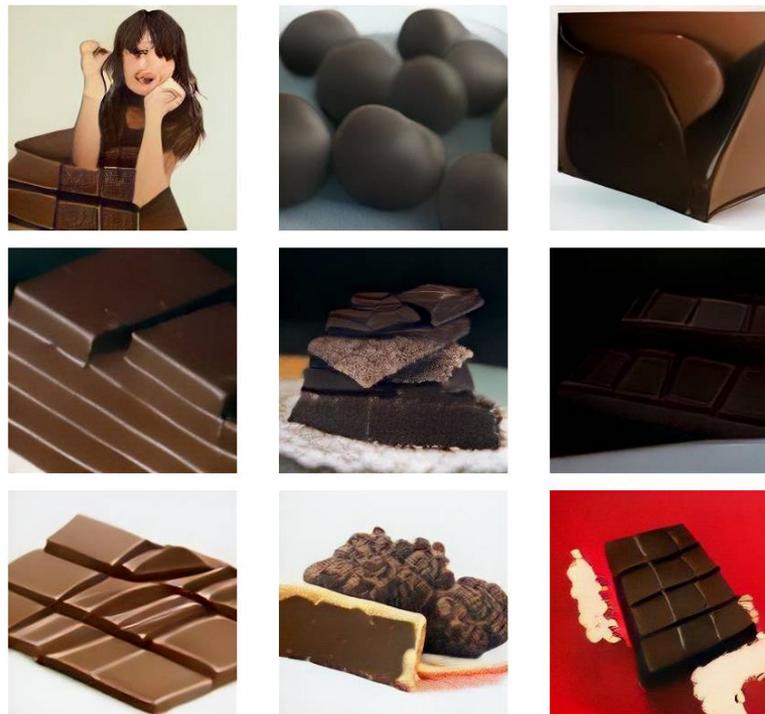
DALL-E mini

Generate images from text

What do you want to see?

I like chocolate

I like chocolate



Again!

AI が生成した画像

DALL-E mini のオンラインデモ

<https://huggingface.co/spaces/dalle-mini/dalle-mini>

文書合成（データの分析）



次の表形式データから製品Aの売り上げと製品Bの売り上げの違いを分析してください

四半期	売上	利益	新規顧客数	製品Aの売上	製品Bの売上
Q1	150	30	200	1000	1500
Q2	200	50	250	1200	1700
Q3	180	40	220	1150	1600
Q4	220	60	300	1300	2000

製品Aの売上推移を見ると、Q1から順に1000、1200、1150、1300となっており、比較的安定しています。

一方、製品Bの売上推移は、Q1から順に1500、1700、1600、2000と上昇しています。

したがって、製品Bの売り上げは製品Aの売り上げよりも毎四半期で高くなっており、製品Bの方が売り上げが伸びていると言えます。

AI の回答



あなた

Python で、うるう年を求めるプログラムを作成してください

ChatGPT

以下にPythonでうるう年を求めるプログラムの例を示します。

```
def is_leap_year(year):  
    # 4で割り切れない年はうるう年ではない  
    if year % 4 != 0:  
        return False  
  
    # 100で割り切れる年はうるう年ではないが、400で割り切れる年はうるう年  
    if year % 100 == 0 and year % 400 != 0:  
        return False  
  
    return True  
  
year = int(input("年を入力してください: "))  
  
if is_leap_year(year):  
    print(year, "年はうるう年です")  
else:  
    print(year, "年はうるう年ではありません")
```

このプログラムでは、`is_leap_year()` 関数を定義しています。この関数は、引数として受け取った年がうるう年かどうかを判定し、結果を返します。`is_leap_year()` 関数がTrueを返す場合、与えられた年はうるう年であると判断します。

プログラムでは、入力された年を `year` 変数で受け取り、`is_leap_year()` 関数に渡して結果を判定します。そして、結果に応じてメッセージを表示します。

このプログラムを実行すると、コンソールに年を入力するプロンプトが表示されるので、任意の年を入力してください。すると、プログラムはその年がうるう年であるかどうかを判定し、結果を表示します。

AI の回答

翻訳（日本語から英語，英語から日本語）

The screenshot shows the DeepL translator interface. At the top, there are two buttons: "テキスト" (Text) and "ドキュメント" (Document). Below these are language selection menus. The source language is set to "日本語 - 自動検出" (Japanese - Auto-detect) and the target language is "英語" (English). The input text is "白い雲と青い空が美しい" (Shiroi kumo to aoi sora ga utsukushi). The output text is "Beautiful white clouds and blue sky". There are icons for voice input/output, a character count of 11/5000, and a star icon for saving the translation. A "フィードバックを送信" (Send feedback) link is visible at the bottom right.

The screenshot shows the DeepL translator interface. At the top, there are two buttons: "テキスト" (Text) and "ドキュメント" (Document). Below these are language selection menus. The source language is "英語" (English) and the target language is "日本語" (Japanese). The input text is "Once upon a midnight dreary, while I pondered, weak and weary, Over many a quaint and curious volume of forgotten lore, |". The output text is "深夜の惨めな時に、私が熟考しながら、弱くて疲れきった、忘れられていた伝説の古風で不思議なボリュームの上に、Shin'ya no mijimena toki ni, watashi ga jukukō shinagara, yowakute tsukare kitta, wasure rarete ita densetsu no kofūde fushigina boryūmu no ue ni,". There are icons for voice input/output, a character count of 120/5000, and a star icon for saving the translation.

翻訳（ファイルの使用）

ワード、パワーポイントファイルの中から、
文字のデータを取り出すことも当たり前

ue-2. C++ プロジェクトの新規作成 (Unreal Engine 演習シリーズ)

<https://www.kkaneko.jp/cc/ue/index.html>

金子邦彦



① Unreal エディタのメインメニューで、「編集」、「プラグイン」と操作



② 「Scripting」を選び、「Python Editor Script Plugin」の「有効」をチェック



ue-2. Create a new C ++ project (Unreal Engine exercise series)

Kunihiko Kaneko

<https://www.kkaneko.jp/cc/ue/index.html>

1 In the main menu of the Unreal Editor, operate "Edit" and "Plug-in"

2 Select "Scripting" and check "Enable" in "Python Editor Script Plugin"

3 Confirm the message and "Yes"

- Create Project Visual Studio code is generated
- Visual Studio project is open

翻訳結果

元のパワーポイントファイル

要約



あなた

次を簡潔な箇条書きで100文字以内で要約してください：
日本国憲法前文

日本国民は、正当に選挙された国会における代表者を通じて行動し、われらとわれらの子孫のために、諸国民との協和による成果と、わが国全土にわたって自由のもたらす恵沢を確保し、政府の行為によって再び戦争の惨禍が起ることのないやうにすることを決意し、ここに主権が国民に存することを宣言し、この憲法を確定する。そもそも国政は、国民の厳粛な信託によるものであつて、その権威は国民に由来し、その権力は国民の代表者がこれを行使し、その福利は国民がこれを享受する。これは人類普遍の原理であり、この憲法は、かかる原理に基くものである。われらは、これに反する一切の憲法、法令及び詔勅を排除する。

日本国民は、恒久の平和を念願し、人間相互の関係を支配する崇高な理想を深く自覚するのであつて、平和を愛する諸国民の公正と信義に信頼して、われらの安全と生存を保持しようと決意した。われらは、平和を維持し、専制と隷従、圧迫と偏狭を地上から永遠に除去しようと努めてゐる国際社会において、名誉ある地位を占めたいと思ふ。われらは、全世界の国民が、ひとしく恐怖と欠乏から免かれ、平和のうちに生存する権利を有することを確認する。

われらは、いづれの国家も、自国のことのみに専念して他国を無視してはならないのであつて、政治道徳の法則は、普遍的なものであり、この法則に従ふことは、自国の主権を維持し、他国と対等関係に立たうとする各国の責務であると信ずる。

日本国民は、国家の名誉にかけ、全力をあげてこの崇高な理想と目的を達成することを誓ふ。

ChatGPT

日本国憲法の前文は、国民が選ばれた代表者を通じて行動し、戦争を防ぐために平和を維持し、他国との協調を促進し、自由を保護することを宣言している。国政は国民の信託に基づき、その権力と福利は国民に由来する。日本国民は恒久の平和を願い、他国を無視せずに自国の利益と対等の関係を維持する責務を持つと信じている。国家の名誉のために、この理想と目的を達成するために全力を尽くすことを誓う。

AI の回答

ここまでのまとめ

自然言語処理は、人間が普段使う言語（日本語、英語など）をコンピュータが理解したり、生成する技術

応用例

- **情報検索**：インターネットから情報を検索
- **AIとの対話、AIへの指示**：AIが対話や指示に応じる
- **プログラミング支援**
- **人間の指示による文書の作成や推敲**
- **翻訳**：異なる言語間で翻訳
- **要約**：文章から要点を抜き出す

ディープラーニングによる自然言語処理

- GPT-4 (Generative Pre-trained Transformer 4)
 - 大規模な言語モデル
 - データや指示から、高品質なテキスト生成が可能
 - 文章生成、要約、翻訳、プログラミングなどに応用

Pythonによる日本語の処理

- 単語の区切りがないため対処が必要
 - 形態素解析により単語に分割 (例: MeCab、Janome)
- 敬語などの言語的特徴
- Pythonの自然言語処理ライブラリ
 - NLTK (Natural Language Toolkit)
自然言語処理ライブラリ
テキストの前処理、品詞タグ付け、構文解析などが可能
多言語対応し、日本語の形態素解析にも利用可能
 - spaCy
自然言語処理ライブラリ
固有表現抽出、依存構文解析、ベクトル化などが可能
多言語対応し、日本語モデルも提供されている

OCRと音声合成における自然言語処理の役割

- OCR (Optical Character Recognition)
 - 画像から文字を抽出し、テキストデータに変換するプロセス
 - 文字認識後、自然言語処理技術を適用
 - 言語モデルによる誤認識の修正や文脈の理解
 - 形態素解析や構文解析による構造化
- で活用可能に
- 音声合成
 - テキストから自然な合成音声を生成
 - テキストの言語的特徴の分析
 - 形態素解析や構文解析による韻律の制御
 - 高品質な音声合成には、自然言語処理と音声処理の連携が重要
- OCRと音声合成の実際の応用例
- 紙媒体の情報をデジタル化、名刺管理、車番号認識など
- 音声合成: 音声アシスタント、書籍の読み上げ、ナレーション生成など

今回の授業で得られる満足感

- ① 自然言語処理の基礎から応用までを網羅的に理解
- ② 言語モデルと活用法についての知識習得
- ③ Pythonによる日本語処理について
- ③ OCRや音声合成などの話題の理解