

# GPTでオンライン学習教材を自動生成する試み

喜多 敏博

熊本大学 半導体・デジタル研究教育機構 総合情報学部門 教授システム学分野

- 教師が設定した最終的な学習目標
- 学習者の習熟度レベルの記述

を元に、オンライン小テストと、学習資料を生成することを試んでいます。

## これまで行ってきたワークショップの情報

GSIS-LT 連続WS(第5回,第6回)「GPTでオンライン学習教材を自動生成する」

<https://mdqu.rcis.jp/course/view.php?id=26>

GPTを使用したMoodleクイズ問題の生成

<https://github.com/kita-toshihiro/gpt-moodle-content/blob/main/gpt-moodle1-ws0826.ipynb>

GPTを使用したMoodle小テスト問題（ドラッグ&ドロップテキスト）の生成

<https://github.com/kita-toshihiro/gpt-moodle-content/blob/main/gpt-moodle2-ws0826.ipynb>

## 生成する教材における学習目標を指定する

```
In [1]: goal = "データサイエンス倫理に反しないアンケートを実施することができること"  
# goal = "\(\ \log_2(x-1) = 3 \)" のような、対数を含む方程式の解を求めることができること"
```

## 学習者の状況（既修得事項など）

```
In [2]: cond1 = '''  
- 学習者は、アンケートにおけるインフォームド・コンセントの重要性をわかっている  
- 学習者は、アンケート設計における匿名性と機密性の原則を理解している  
'''
```

OpenAI GPT-3.5を使用

```
In [3]: import openai  
openai.api_key = input()  
model="gpt-3.5-turbo-1106" # output max 4k  
#model="gpt-4-1106-preview" # 128k, output max 4k  
#model="gpt-4-32k"  
max_tokens=4000
```

## まず、ナレッジマップ（案）を生成

(ナレッジマップ：学習目標を達成するための下位目標の階層構造。学習課題分析。)

```
In [4]: pr_kmap1 = f'''
以下の学習目標を達成するのに必要な下位の学習目標(sub goals)の階層構造を表す\
ナレッジマップをJSON形式で出力してください。
学習目標：{goal}
ナレッジマップ：
'''

response = openai.ChatCompletion.create(
    model = model,
    messages=[
        {"role": "user", "content": pr_kmap1},
    ],
)
res_kmap1 = response.choices[0]["message"]["content"].strip()
print(res_kmap1)
with open('kmap1.txt', 'w') as f:
    print(res_kmap1, file=f)
```

```
{
  "学習目標": "データサイエンス倫理に反しないアンケートを実施することができること",
  "下位の学習目標": {
    "データ収集の倫理的側面を理解する": {
      "データ収集の目的と方法に関する倫理的考慮を理解する": {},
      "プライバシーと個人情報保護の重要性を理解する": {}
    },
    "アンケート設計における倫理的考慮を理解する": {
      "バイアスや偏りを排除するための設計原則を理解する": {},
      "参加者の心理・感情に配慮した質問の作成方法を理解する": {},
      "結果の公正かつ正確な表現に関する倫理的考慮を理解する": {}
    },
    "データの適切な取り扱いを理解する": {
      "データの保存・保護に関する法的規制を理解する": {},
      "データの解釈や利用における倫理的考慮を理解する": {}
    }
  }
}
```

## ナレッジマップを仕上げる

以下のすべての生成物はこのナレッジマップを参考にしているため、間違いがあると影響大。  
上記のナレッジマップ案をこの時点で人間がちゃんと仕上げる。

```
In [16]: knowledgemapfile = 'kmap1-modified.txt'
#knowledgemapfile = 'kmap1.txt'

with open(knowledgemapfile, 'r') as ff:
    res_kmap1 = ff.read()
print(res_kmap1)

{
  "学習目標": "データサイエンス倫理に反しないアンケートを実施することができること",
  "下位の学習目標": {
    "1": "データ収集の適切な方法を理解する",
    "2": "プライバシーと倫理に配慮したデータ処理方法を学ぶ",
    "3": "偏りのないアンケートのデザイン方法を学ぶ",
    "4": "誤解や誤解を避けるための質問の設計方法を理解する",
    "5": "調査結果の公正な分析と報告手法を学ぶ"
  }
}
```

## 小テストの生成

1. プレーンテキスト形式で小テスト質問を生成
2. 必要に応じて、問題を手動で修正
3. 小テスト問題のテキストを Moodle XMLに変換
4. Moodle XML のフォーマットの不具合を自動修正

Moodleの問題バンクに直接インポートできる moodle-quest1.xml が作成される

### 1. プレーンテキスト形式で小テスト質問を生成

```
In [6]: prompt1 = f'''
あなたはオンラインテストの開発者です。
下記の学習目標に到達しているかどうかを測定できる 4 択問題を 4 問作成してください。
正解の選択肢も示してください。数式の箇所は MathJax形式で書いてください。
学習目標：{goal}
'''

response = openai.ChatCompletion.create(
    model = model,
    messages=[
        {"role": "user", "content": prompt1},
    ],
)
res1 = response.choices[0]["message"]["content"].strip()
print(res1)
with open('quiz1.txt', 'w') as f:
    print(res1, file=f)
```

1. データサイエンス倫理に反しないアンケートを実施するための最も重要なポイントは何か？
  - a) 偽名を使用すること
  - b) 個人情報の匿名化
  - c) 非公開のデータを収集すること
  - d) データの改ざん

正解：b) 個人情報の匿名化

2. データサイエンス倫理に反しないアンケートを実施する際に避けるべき行動は次のうちどれか？
  - a) 参加者の意思に反してデータを収集すること
  - b) 参加者に報酬を与えること
  - c) 参加者の意見を無視すること
  - d) 偽装したデータを作成すること

正解：d) 偽装したデータを作成すること

3. データサイエンス倫理に反しないアンケートを実施する際に重要な考慮事項は何か？
  - a) 参加者の性別や年齢などの個人情報を収集すること
  - b) アンケートの目的や利用方法を明確にすること
  - c) 参加者が提供したデータを他の目的に使用すること
  - d) データの収集を急ぎ行うこと

正解：b) アンケートの目的や利用方法を明確にすること

4. データサイエンス倫理に反しないアンケートを実施する際に配慮すべき点は何か？
  - a) 参加者のデータを無断で他者と共有すること
  - b) 参加者にデータの使用目的を明確に伝えること
  - c) 偽名でアンケートに回答をさせること
  - d) 参加者のデータを勝手に削除すること

正解：b) 参加者にデータの使用目的を明確に伝えること

### 2. 必要に応じて、問題を手動で修正

```
In [ ]: quizfile = 'quiz1.txt'
# quizfile = 'quiz1-modified.txt'

with open(quizfile, 'r') as f:
    res1 = f.read()
print(res1)
```

### 3. 小テスト問題のテキストを Moodle XMLに変換

```
In [8]: prompt2 = '''
さきほどの問題を一連のデータとして（問題ごとに分けずに） \
以下の例のような Moodle XML 形式で出力してください。
以下の例の <text><![CDATA[xxxx]]</text> や <text><![CDATA[ooo]]</text> \
などの箇所は、そのとおりに書くのではなく、各問題での不正解や正解の選択肢を使って書いてください。
選択肢の句点は削除してください。

<question type="multichoice">
  <name>
    <text>問いの要約</text>
  </name>
  <questiontext format="html">
    <text><![CDATA[問いの文章をここに書く。]]></text>
  </questiontext>
  <shuffleanswers>true</shuffleanswers>
  <answer fraction="0">
    <text><![CDATA[xxxx]]</text>
  </answer>
  <answer fraction="100">
    <text><![CDATA[ooo]]</text>
  </answer>
  <answer fraction="0">
    <text><![CDATA[yyyy]]</text>
  </answer>
  <answer fraction="0">
    <text><![CDATA[zzz]]</text>
  </answer>
  <generalfeedback>
    <text><![CDATA[全般に対するフィードバック]]></text>
  </generalfeedback>
</question>
''' [1:]

response = openai.ChatCompletion.create(
    model = model,
    max_tokens = max_tokens,
    messages=[
        {"role": "user", "content": prompt1},
        {"role": "assistant", "content": res1},
        {"role": "user", "content": prompt2},
    ],
)

res2 = response.choices[0]["message"]["content"].strip()
# print(res2)
```

### 4. Moodle XML のフォーマットの不具合を自動修正

```
In [9]: prompt3 = '''
以下のよう直してください。
・1行目が <quiz> であり、全体が一つのXML形式のデータである
・選択肢番号は削除する
・選択肢文字 (a, b, c など) は削除する
・正解の選択肢の配点は 100 になっている
・問題名 (<name><text> と </text></name> の間) は、問いの文章を10文字程度で要約したもの
・各問題に対して1つだけ、</questiontext> の後に、200文字程度の「全般的なフィードバック」を、
<generalfeedback><text> と </text></generalfeedback> で囲って書く
・「全般的なフィードバック」には、参考となるWikipediaページなどのリンクが、
target= BLANK の指定付きで含まれている
・各問題に対して <shuffleanswers>true</shuffleanswers> を指定する
・[CDATA[ に対応する ]] がちゃんと書かれている
'''[1:]

response = openai.ChatCompletion.create(
    model = model,
    max_tokens = max_tokens,
    messages=[
        {"role": "user", "content": prompt1},
        {"role": "assistant", "content": res1},
        {"role": "user", "content": prompt2},
        {"role": "assistant", "content": res2},
        {"role": "user", "content": prompt3},
    ],
)
res3 = response.choices[0]["message"]["content"].strip()
print(res3)
res3 = res3.replace("```xml", '').replace("```", '')
with open('moodle-quest1.xml', 'w') as f:
    print(res3, file=f)
```

```

` ``xml
<quiz>
  <question type="multichoice">
    <name>
      <text>倫理的なアンケート</text>
    </name>
    <questiontext format="html">
      <text><![CDATA[データサイエンス倫理に反しないアンケートを実施するための最も重要なポイントは何
か? ]]></text>
    </questiontext>
    <shuffleanswers>>true</shuffleanswers>
    <answer fraction="0">
      <text><![CDATA[偽名を使用すること]]></text>
    </answer>
    <answer fraction="100">
      <text><![CDATA[個人情報の匿名化]]></text>
    </answer>
    <answer fraction="0">
      <text><![CDATA[非公開のデータを収集すること]]></text>
    </answer>
    <answer fraction="0">
      <text><![CDATA[データの改ざん]]></text>
    </answer>
    <generalfeedback>
      <text><![CDATA[データの倫理に関する重要なポイントです。参考: <a href="https://en.wikipedi
a.org/wiki/Data_ethics" target="_blank">Data Ethics</a>]]></text>
    </generalfeedback>
  </question>

  <question type="multichoice">
    <name>
      <text>アンケートの避けるべき行動</text>
    </name>
    <questiontext format="html">
      <text><![CDATA[データサイエンス倫理に反しないアンケートを実施する際に避けるべき行動は次のうちど
れか? ]]></text>
    </questiontext>
    <shuffleanswers>>true</shuffleanswers>
    <answer fraction="0">
      <text><![CDATA[参加者の意思に反してデータを収集すること]]></text>
    </answer>
    <answer fraction="0">
      <text><![CDATA[参加者に報酬を与えること]]></text>
    </answer>
    <answer fraction="0">
      <text><![CDATA[参加者の意見を無視すること]]></text>
    </answer>
    <answer fraction="100">
      <text><![CDATA[偽装したデータを作成すること]]></text>
    </answer>
    <generalfeedback>
      <text><![CDATA[データの不正な操作は回避すべきです。参考: <a href="https://en.wikipedia.or
g/wiki/Online_ethics" target="_blank">Online Ethics</a>]]></text>
    </generalfeedback>
  </question>

  <question type="multichoice">
    <name>
      <text>アンケートの重要な考慮事項</text>
    </name>
    <questiontext format="html">
      <text><![CDATA[データサイエンス倫理に反しないアンケートを実施する際に重要な考慮事項は何か? ]]>
</text>
    </questiontext>
    <shuffleanswers>>true</shuffleanswers>
    <answer fraction="0">
      <text><![CDATA[参加者の性別や年齢などの個人情報を収集すること]]></text>
    </answer>
    <answer fraction="100">
      <text><![CDATA[アンケートの目的や利用方法を明確にすること]]></text>
    </answer>
    <answer fraction="0">
      <text><![CDATA[参加者が提供したデータを他の目的に使用すること]]></text>
    </answer>

```

```

<answer fraction="0">
  <text><![CDATA[データの収集を急ぎ行うこと]]></text>
</answer>
<generalfeedback>
  <text><![CDATA[アンケートの目的と利用方法の明確化が重要です。参考: <a href="https://en.wikipedia.org/wiki/Ethical_code" target="_blank">Ethical Code</a>]]></text>
</generalfeedback>
</question>

<question type="multichoice">
  <name>
    <text>アンケートの配慮すべき点</text>
  </name>
  <questiontext format="html">
    <text><![CDATA[データサイエンス倫理に反しないアンケートを実施する際に配慮すべき点は何か?]]></text>
  </questiontext>
  <shuffleanswers>true</shuffleanswers>
  <answer fraction="100">
    <text><![CDATA[参加者にデータの使用目的を明確に伝えること]]></text>
  </answer>
  <answer fraction="0">
    <text><![CDATA[参加者のデータを無断で他者と共有すること]]></text>
  </answer>
  <answer fraction="0">
    <text><![CDATA[偽名でアンケートに回答をさせること]]></text>
  </answer>
  <answer fraction="0">
    <text><![CDATA[参加者のデータを勝手に削除すること]]></text>
  </answer>
  <generalfeedback>
    <text><![CDATA[データの使用目的の明確化がポイントです。参考: <a href="https://en.wikipedia.org/wiki/Privacy_policy" target="_blank">Privacy Policy</a>]]></text>
  </generalfeedback>
</question>
</quiz>

```

## 上記XMLをインポートした結果（Moodle上での表示）

The screenshot shows the Moodle interface for a quiz titled '小テスト1'. The page is in preview mode. At the top, there is a navigation bar with 'ds-ethics / 小テスト1 / プレビュー'. Below this, the quiz title '小テスト1' is displayed with a checkmark icon. A secondary navigation bar contains '小テスト', '設定', '問題', '受験結果', '問題バンク', and 'さらに▼'. The main content area features a '戻る' button and two question cards. Each card shows the question number, status '未解答', and maximum score '最大評点 1.00'. Question 1 asks for the most important point for implementing an ethical survey, with options: a. Using pseudonyms, b. Anonymizing personal information, c. Collecting non-public data, and d. Not changing data. Question 2 asks for actions to avoid when implementing an ethical survey, with options: a. Creating fake data, b. Ignoring participant opinions, c. Collecting data against participant intent, and d. Rewarding participants.

## 学習資料の生成

1. 学習項目と習得済項目をJSON形式で一覧表示する
2. 学習項目それぞれに説明文を作成する
3. JSONから学習項目を1つずつ抽出し、説明文に小テストを追加する
4. HTML形式で学習資料を作成する
5. 資料をPDFに変換する（おまけ）

学習用の資料 out.html が作成される

### 1. 学習項目と習得済項目をJSON形式で一覧表示する

```
In [10]: prompt11 = f'''
あなたはeラーニング教材の開発者です。今回開発するeラーニング教材の学習目標は、
{goal}
です。
ナレッジマップを参考にして、eラーニング教材が含むべき学習項目をいくつか（5個程度）列挙してください。
また、下記の「教材の対象者（学習者）の状況」に基づいて、学習者が習得済だと判断できる項目も列挙してください。
学習項目を決める際には、習得済項目は除外し、学習目標到達までのギャップを埋めるのに必要な項目を挙げてください。

出力形式は、
{{
  "学習項目": [
    "学習項目1. ....",
    "学習項目2. ....",
    "学習項目3. ....",
  ],
  "習得済項目": [
    "習得済項目1. ....",
    "習得済項目2. ....",
  ]
}}
のようなJSON形式にしてください。

教材の対象者（学習者）の状況：
{cond1}
'''

response = openai.ChatCompletion.create(
    model = model,
    max_tokens = max_tokens,
    messages=[
        {"role": "user", "content": pr_kmap1},
        {"role": "assistant", "content": res_kmap1},
        {"role": "user", "content": prompt11},
    ],
)

res11 = response.choices[0]["message"]["content"].strip()
print(res11)

{
  "学習項目": [
    "データ収集の適切な方法について学ぶ",
    "プライバシーと倫理に配慮したデータ処理方法を理解する",
    "偏りのないアンケートのデザイン方法を学ぶ",
    "誤解や誤解を避けるための質問の設計方法を理解する",
    "調査結果の公正な分析と報告手法を学ぶ"
  ],
  "習得済項目": []
}
```

### 2. 学習項目それぞれに説明文を作成する



```

In [11]: prompt12 = '''
学習項目の文章はそのまま、それぞれの項目に対して、学習者に向けて、少し長めの説明文を書いてください。
高校生でもわかるようなやさしい説明にしてください。
説明文は、ナレッジマップや、最終の学習目標である
「{goal}」
も踏まえつつ、具体的な例示や、学習項目の内容について理解を問う表現も含めるなどして、具体的に書いてください。
習得済と思われることについては説明せず、\
最終の学習目標に到達するために学習者に不足していることを説明してください。
出力は以下のようなJSON形式にしてください。"学習項目" と "説明文" は、\
それぞれ上から順番に対応しており同じ個数です。
{{
  "学習項目": [
    "学習項目1. ....",
    "学習項目2. ....",
    "学習項目3. ....",
  ],
  "説明文": [
    "1. ....",
    "2. ....",
    "3. ....",
  ]
}}
'''

response = openai.ChatCompletion.create(
  model = model,
  messages=[
    {"role": "user", "content": pr_kmap1},
    {"role": "assistant", "content": res_kmap1},
    {"role": "user", "content": prompt11},
    {"role": "assistant", "content": res11},
    {"role": "user", "content": prompt12},
  ],
)

res12 = response.choices[0]["message"]["content"].strip()
print(res12)

{
  "学習項目": [
    "データ収集の適切な方法について学ぶ",
    "プライバシーと倫理に配慮したデータ処理方法を理解する",
    "偏りのないアンケートのデザイン方法を学ぶ",
    "誤解や誤解を避けるための質問の設計方法を理解する",
    "調査結果の公正な分析と報告手法を学ぶ"
  ],
  "説明文": [
    "1. データ収集の適切な方法について学ぶことは、例えば、インターネット上の情報を使わずに信頼性の高いデータを集める方法や、データ収集における個人情報や機密情報の取り扱いを学びます。",
    "2. プライバシーや倫理に配慮したデータ処理方法を理解することは、例えば、個人情報を匿名化する手法や、データセキュリティを確保する方法について学びます。これによって、アンケートの結果が個人を特定することがないように配慮できます。",
    "3. 偏りのないアンケートのデザイン方法を学ぶことは、例えば、ランダムサンプリングの方法や、適切なバイアスを排除するためのテクニックを学びます。これにより、客観的かつ公平なアンケートを作成できます。",
    "4. 誤解や誤解を避けるための質問の設計方法を理解することは、質問の明確さや中立性を確保する方法について学びます。これによって、回答者が誤解されことなく正確な情報を得ることができます。",
    "5. 調査結果の公正な分析と報告手法を学ぶことは、統計的手法や結果の透明性を求める理由について学びます。これによって、アンケート結果を客観的に分析し、公平かつ正確に報告するスキルを身につけます。"
  ]
}

```

### 3. JSONから学習項目を1つずつ抽出し、説明文に小テストを追加する

```
In [12]: import json
res12esc = res12.replace('\\', '\\\\') # MathJax数式での "Invalid \escape" エラーを避ける
res12esc = res12esc.replace("`json", '').replace("`", '')
topics = json.loads(res12esc)
i = 0
res13 = ""
for li in topics['学習項目']:
    print(i)
    de = topics['説明文'][i]
    str = f'<h2 style="color:darkgreen; margin-top:2em;">{li}</h2><p>{de}</p>'
    i += 1
    prompt13s = f'''
{str}

{i}番目の学習項目についての上記のHTML形式の記述を以下の方針に従って書き直してください。

・HTML形式で出力する。
・<h2></h2>付近は書き換えずそのままに。
・Wikipediaの参考ページなどのリンクを記載する。
・記述のうち、要点と思われる箇所は、<strong></strong>で囲う。
・記述内容に関連する小テストを3問作成して追記する。
・多肢選択式の小テストの場合は、必ず「上記に正解はない」の選択肢を設ける。
・<input>は用いない。
・数式が書かれた箇所については、MathJax形式で書く。\\
ただし、<script></script>などのJavaScriptコードは追記しない。
'''

    #print(prompt13s)
    response = openai.ChatCompletion.create(
        model = model,
        max_tokens = max_tokens,
        messages=[
            {"role": "user", "content": pr_kmap1},
            {"role": "assistant", "content": res_kmap1},
            {"role": "user", "content": prompt11},
            {"role": "assistant", "content": res11},
            {"role": "user", "content": prompt13s},
        ],
    )
    res13s = response.choices[0]["message"]["content"].strip()
    #print(res13s)
    res13 += res13s
```

```
0
1
2
3
4
```

## 4. HTML形式で学習資料を作成する

(HTMLヘッダに色々としているのは、この次のPDF変換の際の文字化け防止と、MathJax数式表示のため)

```

In [13]: htmlhead = '''
<html lang="en"><head>
<style>
body {
  font-size: 20px;
  font-family: "IPAexMincho", "IPAex明朝";
}
</style>
<meta charset="utf-8"></meta>
<script type="text/x-mathjax-config">
MathJax.Hub.Config({
  TeX: {extensions: ["mhchem.js"]},
  tex2jax: {
    processEscapes: true
  }
});
</script>
<script type="text/javascript" async
  src="https://cdnjs.cloudflare.com/ajax/libs/mathjax/2.7.1/MathJax.js?config=TeX-MML-A
</script>
</head><body>
...
htmlfoot = '</body></html>'

outs = htmlhead + res13 + htmlfoot
outs = outs.replace("```json", '').replace("```html", '').replace("```", '')
with open('out.html', 'w') as f:
  print(outs, file=f)

```

## (おまけ) 資料をPDFに変換する

```

In [14]: import pdfkit
options = {
  'quiet': '',
  'javascript-delay' : '5000',
  'page-size': 'A4',
  'margin-top': '0.75in',
  'margin-right': '0.75in',
  'margin-bottom': '0.75in',
  'margin-left': '0.75in',
  'disable-smart-shrinking': '',
  'dpi': '400',
  'minimum-font-size': 18,
}
pdfkit.from_file('out.html', 'out.pdf', options=options)

```

Out[14]: True