

生成 AI による情報プレースメントテストのための 設問生成の実践と評価

中鉢直宏¹ 久野靖² 角田博保² 中山泰一²

概要: 本稿は、過去の IPT 実施結果を踏まえ、2025 年度の大学入学者向けに改訂された「情報 I」科目に準拠した IPT2025 の作問プロセスの改善に取り組むものである。IPT2025 の作問は、情報教育知識体系である GEBOK (General Education Body of Knowledge) に準拠し、GEBOK2027 改定を見据えた設問構成を目指し、従来の IPT 作問作業においては、設問の質や作問コスト、レビュー不足、難易度調整といった課題を ChatGPT を用いることで課題解決を目指す。ChatGPT による生成設問の質について評価するため、生成 AI により背苦悶された設問に対して専門家による評価を実施し、生成 AI による作問が人間作問と比較してどの程度実用性があるかを検証し、生成 AI による作問は一部効果が認められたが、選択肢の生成において質の低下が見られたことが分かった。

キーワード: 一般情報教育, IPT, 情報プレースメントテスト, 生成 AI, ChatGPT

Practice and Evaluation of Question Generation for Information Placement Test by Generative AI

CHUBACHI NAOHIRO^{†1} KUNO YASUSHI^{†2}
KAKUDA HIROYASU^{†2} NAKAYAMA YASUICHI^{†2}

1. はじめに

現代社会において情報環境と関連技術は常に進化が続いている。このような背景を受けて、初中等教育段階における情報教育が行われるようになり、高校は教科「情報」が必修化された。このような変化からこれから大学の入学時に新入生が持っている情報知識とスキルも大きく変化すると予測される。そのため、情報プレースメントテスト（以下：IPT）の実施の重要性が増している。IPT は、情報処理学会一般情報教育委員会（以下：本委員会）が過去 2018 年度（以下、IPT2018）、2020 年度（以下、IPT2020）と実施してきたプロジェクトで、大学にどのような情報知識を持つ学生が入学するかを調査するために複数の大学にて情報知識に関するテストを実施してきており、新入生の持つ知識空間の状態遷移を知ることが狙いとして調査目的で実施されてきた[1]。そのため、大学入学時に高校生がどのような情報知識を有しているかを、高校で習得される内容だけを調査対象とするのではなく、大学の科目として広く設けられている一般情報教育という領域に関しての内容まで含んだ範囲の設問を作成し、テストを実施してきた。一般情報教育は、高度情報化社会において中核を担う大学生に対し、文系理系を問わず情報に関する基礎理論や応用技術を身に付けさせることを目的とした教育分野である[2]。

これまで実施された IPT の結果は、設問の正解率などから大学入学時の学生の情報の知識空間の把握を目指してきた。

令和 4 年度より、新しい高等学校学習指導要領に基づき、高等学校情報科においては共通必修科目「情報 I」が新設される[3]、これにより 2025 年度の大学新入生が持っている情報に関する知識とスキルは大きく変化すると考えられている。

現在、高校の教科「情報」の必修化の影響が出てくるであろう 2025 年度の入学者に向けて、IPT2025 を実施するために現在は IPT2020 の設問について見直し、IPT2025 のための新しい設問を用意するために準備に取り組んでいる。

2. IPT2025 の目的と作問作業

IPT2025 の実施主体である本委員会は、これまで一般情報教育の知識体系を示した GEBOK (General Education Body of Knowledge) の策定を行ってきた。現在は、GEBOK2017.1 が公表されている[4]。現在、次期バージョンである GEBOK2027 のために高校生の情報に関する知識がどのくらい習得されているか把握する必要がある。そのため、IPT2020 までの設問は GEBOK で示されている領域を基にしているものの、その時々ニーズに合わせるため、GEBOK とは一致しない領域単位で出題されていた。しかし IPT2025 では、明確に GEBOK2027 に向けた調査を目的としているためこれまでの出題範囲を見直すこととなった。

表 1 は、大学の情報環境について取り上げ「科目ガイド

¹ 高崎商科大学
Takasaki University of Commerce

² 電気通信大学
The University of Electro-Communications

ンス」を除いた GEBOK2017.1 と IPT2020 の設問領域を示している。IPT2020 は、一般情報教育委員会のメンバーにより GEBOK を参考に 10 領域が設定された。しかし、IPT2020 で行われた作問の 10 領域は GEBOK2017.1 の領域と多少異なっている[5]。現在検討されている IPT2025 の作問については、GEBOK2027 の改定を見据え、GEBOK2017 の領域にベースとして実施することを計画している。

表 1 IPT2020 設問の領域と GEBOK2017.1 の領域

GEBOK2017の領域	IPT2020の設問領域
GE-ICO 情報とコミュニケーション	情報とコミュニケーション
GE-DIG 情報のデジタル化	情報のデジタル化
GE-CEO コンピューティングの要素と構成	コンピューティングの要素と構成
GE-ALP アルゴリズムとプログラミング	N/A
GE-SIM モデル化とシミュレーション	N/A
GE-DMO データベースとデータモデリング	データモデルとデータベース
GE-AID 人工知能 (AI) とデータ科学	N/A
GE-INW 情報ネットワーク	情報ネットワーク
GE-INS 社会と情報システム	社会と情報システム
GE-ISE 情報セキュリティ	情報倫理と情報セキュリティ
GE-IET 情報倫理	
GE-AIL アカデミック ICT リテラシー	アカデミックICTリテラシー
N/A	問題解決技法
N/A	メディアとコンピュータの歴史と未来

(N/A: 該当する領域がない)

2.1 IPT2020 の作問方法

IPT2020 の作問は、IPT2020 の 10 領域に対し、一般情報教育委員会のメンバー 1 人に対し 1 つの領域を作問担当者として割り当てられて行われた。各担当者は、IPT の作問においては次の IPT の基本方針 (図 1) に従い、各領域に必要な情報知識について作問を行った。各領域には 20 問の設問数のノルマが設定された。作成された設問については、領域ごとに別の作問者が相互レビューを行い、作問内容の精査が行われた。

-
- ①.大学1年生に必要な知識を問うテスト
 - ②.合否を目的とするテストにしない
 - ③.質問はできるだけシンプルに行う
 - ④.テスト結果を以下の目的で使用
 - 1.情報に関する知識調査
 - 2.一般情報教育の授業評価
 - 3.結果にあわせた授業計画の変更
 - 4.結果によるクラス組織

図 1 IPT の基本方針

2.2 IPT の設問形式

IPT の設問は 5 肢択一形式となっている。ただし、5 肢択一の選択肢は、1 つの「正解」、3 つの「誤答」、1 つの「わからない」で構成される。設問で問われている知識に対し、受験者がそれを知っているかどうかを確認するため、誤答と区別する目的で『わからない』の選択肢を設けている。このため作問担当者は、実質 4 肢択一を考えるとなる。

また、設問ルールとして、受験者が混乱しないように設問は「間違っているものを選べ」という形での設問は作問しないようにした。設問文の内容は原則として正しいものを選ぶ形式に誘導し、必ず選択肢の中から正解を 1 つ選ぶようにする必要がある。このような作問作業を行っていくうえで様々な作問上の課題が発見されることとなった。

2.3 IPT における作問の課題

2020 年度の IPT 作問方法で判明した課題は以下の通りである [7]

- ・ 設問の質
- ・ 作問コスト
- ・ 選択肢の誤答作成
- ・ 全体のレビュー
- ・ 難易度

設問の質とは、設問内容の正確さや傾向に個人差が生じてしまった問題である。設問や選択肢において、文面から正解が推測できてしまうものがあり、その設問では知識を図りかねるような設問が確認され、設問の質に関する問題が起きてしまっていた。また、領域ごとに担当者が異なるため設問の統一感や表現方法などの違いなどの問題があった。

作問コストとは、作問作業は作問者への大きな負担の問題である。本来の業務に加え 20 問という作問量は作問担当者に大きな負担となった。作問者にとってかなりの時間を要することになってしまい、実際にどうしても 20 問作問できない領域が生じたり、作問コストを削減が求められた。

選択肢の誤答作成とは、作問者が間違った選択肢 (以下誤答) を作成するときの問題である。実際に設問作成時に作問者にとって大きな負担となったのは、選択肢の誤答を作る作業である。作問者は設問ごとに間違っている選択肢を 3 つ用意する必要があった。一人当たり 20 問、計 60 問の間違えの設問を作成することは作問者にとって大きな負担となったと報告された。そのため、誤答作成の省力化が必要とされた。

こうした問題の影響として、**全体のレビュー**が不足する事態が生じた。本来、全体の設問について整合性を確保するため、200 問にわたる設問を一括でレビューし、統一感や領域間の重複問題を調整する必要があった。問題を精査するだけでなく、問題の統一感や領域を超えて同じような問題が存在するなどの領域間を調整の作業が必要とされた。しかし、は横断的に全体をレビューするには先のコストの問題もあり、領域担当者に事情を伝え、領域間で調整してもらいなどの、全体レビューの担当者だけでなく、個々の作問担当者にも影響があるため、十分な全体レビューは行えたとはいえない環境であった。

難易度については、各担当者が想定する新入生の知識レベルにばらつきがあったため、難易度の調整が必要であった。しかし、全体レビューが不足していたこともあり、作問時に難易度の差を解消できなかった。また、難易度自体の予測が困難であり、実施後に確認すると難しすぎる設問や簡単すぎる設問があり、識別力の低い設問も存在した。

3. 生成 AI (Generative AI) による作問

3.1 生成 AI による作問の可能性

既存の課題を解決すべくこれまで生成 AI を使った新しい作問方法検討することとなった。生成 AI により作問を行うことで、その作問の質の一定化、コストの削減、難易度への配慮による課題解決を目指した。そして筆者は、情報分野において ChatGPT による指定したキーワードから設問の作成、既存の設問のための解答の選択肢生成を行い、ChatGPT を使用した IPT のための作問支援についてその可能性を示している。

3.2 生成 AI サービスの登場と教育利用について

ChatGPT (Chat Generative Pre-trained Transformer) (a)は OpenAI 社が提供する生成 AI サービスとして 2022 年 11 月に登場した。自然言語を使って AI とのやり取りし、文章やプログラムのコードなどを生成することができるようになった。それ以来、生成 AI を様々なサービスが誕生し、生成 AI のモデルの進化によりより精度が高く、自然な生成が行えるようになった。

生成 AI は、様々な分野で人間の作業に関して手助けとなると注目され、文科省は「教員による教材開発や、効果的・効率的な大学事務の運営等に利活用することも考えられる。」とあり教職員の生成 AI の利活用の可能性を取り上げている[5]。そのため生成 AI を使った作問についてもこれから積極的に行われると予想されており、本研究の取り組みはその一助になると考える。

3.3 リサーチクエスト

設問作成に一般情報教育の教科書と生成 AI を使用することにより解消することを目指すこととした。そのために必要な生成 AI による作問についてその手法をと事例を報告してきた。また生成 AI のバージョンがアップ利他ことにより今までの生成時の課題が解決された質が向上したことが可能性される。そこで本研究では、生成 AI により作問された設問に対して専門家による評価を行いどれくらいの質の作問ができたか検討する。

リサーチクエストは以下のようである。

1. 人間と生成 AI が作成した設問と質と差がない
2. 従来の作成方法と比較してコスト削減ができるか

a) OpenAI : ChatGPT, (https://chat.openai.com/)

3. GEBOK に従った作問できるか

3.4 生成 AI による作問方法

2023 年に ChatGPT に Advanced Data Analysis 機能が追加されてから ChatGPT のサービスでできることが飛躍的に増えた。登場時は、Code Interpreter と称された Advanced Data Analysis 機能は、ファイルをアップロードし、そのファイルを分析することによって、アップロードされたデータに基づいて様々な生成が行えるようになった。プログラムソースの解析や画像認識、そして、データ分析においても、資料からデータを抽出して分析できるようになった。そこで作問生成に、ChatGPT でのテキストデータをアップロードして、その内容をもとに回答を生成させることが出来る。そこで、一般情報教育の教科書の内容を基に設問を作成することを試みた。今回は実際に章ごとのテキストファイルを読み込ませて作問を行った。

以前 ChatGPT の GPT-4 (以下 GPT-4) を使用して作問を行って、以下の課題が見つかった[]。しかし今回使用した ChatGPT の GPT-4o (Generative Pre-trained Transformer 4 Omni) (以下 GPT-4o) 2024 年 5 月 13 日に OpenAI によって発表され、テキスト、画像、音声などのマルチモーダル入力に対応し、従来の GPT-4 よりも大幅に機能が強化されたおり、これまでの作問作業の課題が解消されていた。

- ChatGPT の問題
 - 作問に様々な制限がある
 - ChatGPT の動作が安定しない。
- 誤答が短くなってしまう傾向
- 設問が誤答を選ばせる問題になる傾向

GPT4 では、読み込ませるテキストファイルの容量に関して見えない制限があった一度に読み込ませるテキストのデータ量が多いと、最後までファイルを読み込んでいない現象が頻繁に起きた。しかし GPT4o を使用して行った今回は比較的安定し、テキストの読み込みが途中で終わるという現象はほとんど起きなかった。テキストデータを細分化し、細分化したテキストデータから数問の設問作成を行っていた作問作業についても、GPT-4o では、以前よりも多くのテキストデータを読み込ませて安定した出力を得ることができ、まとめて作問の数も 3 つが限界だったものも、今回は教科書の項ごとに 10 問をまとめて作問しても問題なく生成された。

また、以前は、テキストデータに基づいて 4 択問題を作成させると、正解の選択肢が他の誤答の選択肢よりも文字数が多くなってしまいう傾向があることがあり、正答が推測されやすいという問題があったが、これはプロンプトの改良と GPT-4o の性能の向上により、読み込ませる元の文章の性質にもよるが是正されていた。ChatGPT の利用制限の

緩和など大量の作問に耐えうる環境が整ってきたように思える。

4. ITP2024 のための生成 AI による作問方法

4.1 GPTs について

今回の作問も先行研究と同様に GPTs の機能を使って行う。GPTs は、2023 年 11 月に発表された機能であり、特定の目的のために「カスタムされた ChatGPT」(b) (以下、GPT) を作成することが出来る機能である[8]。GPT の作成は、ChatGPT Plus の有料のプランで使用できるものである。MyGPT と呼ばれる GPT を共有することもでき、作問作業を手分けして行えるために採用した。

従来のチャット型のインターフェースだと本来目的に合わせた回答を得るためにはプロンプトを介して生成 AI と会話を行いながら、その振る舞いを生成 AI に指示する必要があるが、GPT の特徴として、予め生成 AI に必要なプロンプトの項目を目的に合わせて記述しておくことができる。MyGPT として登録すれば、最初から目的に合わせた使い方を事前の指示になしで生成させることができるようになる。

今回のように同じ環境で生成結果を得たい場合は、GPT 機能を使用し、MyGPT から生成させることができる。これにより同様の作業が安定して実行することが出来るようになった。しかし、事前のプロンプトが共通で入力したデータが同じであっても、生成 AI の特性上、生成の結果が同様になることは必ずしも一致しないため出力結果の再現性は難しいことには注意が必要である。実際の作問作業では、共有機能により共通の MyGPT を使って作問作業を分担することも視野に入れている。

4.2 MyGPT の作成手順

MyGPT の作成は、ChatGPT の GPT Builder の機能を使用する。GPT Builder では、以下の 8 つの設定項目を記述する必要がある。

1. 「Name (GPT の名前)」
2. 「GPT のアイコン用イメージデータ)」
3. 「Description (短い GPT に関する説明)」
4. 「Instructions (何をするか、どのようにふるまうのか、何を避けるのか等を具体的に指示する)」
5. 「Conversation starters (Chat 開始時の基本会話)」
6. 「Knowledge (ファイルアップロード)」
7. 「Capabilities (追加機能)」
8. 「Actions (他の Web サービスとの連携)」

b) OpenAI はカスタムされた ChatGPT のことを GPT と名称している。

9.

これらの設定項目の記述方法は、Create と Configure の 2 つの方法があり用意されている。

図 4 の GPT Builder の Create の操作画面のように 2 つのペインに分かれており、左側のペインで、GPT Builder が t 投げかけられる質問に答えることで GPT に必要な設定項目の内容を作成することができる。GPT Builder の質問に答えていくと GPT の設定項目がその会話の内容に従って生成をされ記述されていく。また、右のペインの Preview では会話で作成された設定項目が反映されたプロンプトを試すことができ、改良点があれば、左のペインで会話を行い、GPT の設定項目が更新することができる。その設定項目の更新内容は常に右のペインにあるに反映される。Create は、会話と確認作業を複数繰り返すことで設定項目を更新しながら目的の GPT を作成させていく方法である。

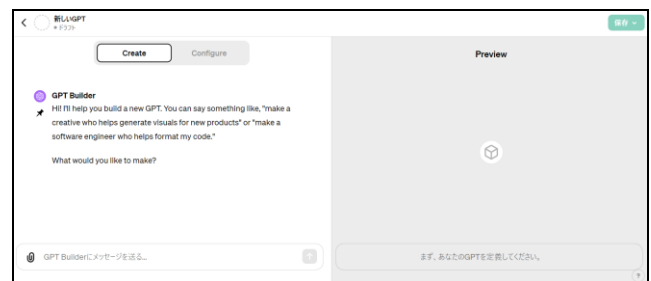


図 2 GPT Builder (Create 画面)



図 3 GPT Builder (Configure 画面)

次に Configure 画面を図 5 に示す。Configure は、Create と同様 2 つのペインに分かれているが左のペインで直接の生成 AI への設定を記述する方法である。

Configure を使用して GPT を作成するときは、Configure の設定画面の左側に直接設定を記述する。プロンプトを作成することに慣れている場合は Configure による項目設定が便利である。

Create の会話の結果も Configure の設定画面に切り替えると左側のペインに会話の内容が反映された設定が記述されている。

いずれかの方法により作成された GPT は MyGPT として保存することができるため、継続したプロンプト開発が行える。

4.3 設問生成のため MyGPT の作成

作成手順は以下のとおりである。まず、Create を使用し全体の設定項目の概要を作成した。作成したい GPT の目的を伝えると GPT Builder から名前とアイコンが生成される。次に、GPT Builder は会話により Instruction の設定項目を生成するやり取りを行う。ここでは以下の指示を会話形式で行った。

- ・ 情報に関する 4 択の選択肢を作ること
- ・ 正解の選択肢と間違った選択肢が同じ文字数になるようにすること
- ・ 設問を作成するときは、設問文は正しい選択肢を選ばせることを伝えること。

このほかには、応答時の口調や振る舞いについて質問される。これらの指示が設定項目に直接記述されるのではなく、GPT Builder との会話の中で設定項目を生成してから反映される。たとえば、今回のやり取りでは、「指導とアドバイスを重視し、ユーザーの要望に応じた適切な問題を提案します。」を GPT Builder が生成し付け加えている。このようなやり取りを経て、一通り会話で作成した後 Preview を使用し、さらなる生成結果の検討を行う。

その後、Configure の操作画面を使用し調整などを行った。実際には、選択肢に、1 から 4 の選択肢の番号を付与したり、設問に否定的な内容を含まないようにするなど設問のルールや出力について、後から直接設定項目に追加した。設定として、インターネットを使った情報のためのウェブ検索機能と高度なファイルの書き出しなどを行うためにコードインタープリターとデータ分析機能を有効にしている。

今回作成して作問に使用した GPT の設定は付録を示す。

5. 生成 AI による作問評価

5.1 評価の目的

今回の作問評価の目的は、生成 AI の作問に対してどのような評価を行う必要があるのか検討することである。また、現状の生成 AI による作問方法で作問された設問が、どれくらい設問として成立しているかについて第三者による評価で確認することを目的としている。そのため、大学における情報の専門家に協力を仰ぎ、設問に対しての評価を行った。

5.2 評価方法

評価方法の概要は以下に示す (図 3)。

IPT2025 の領域は「情報とコミュニケーション」を選んだ。生成方法は一般情報教育の教科書の情報とコミュニケ

ーションの章のテキストデータを項ごとに分け、作成した MyGPT にアップロードし 1 つの項のデータに対し、10 問の設問を生成させた。この章は 3 つ項からなるため 30 問作問することができた。出力された作問の評価時に統一感を出すためにフォーマット等を手直しを行ったものを使用した。

今回の評価では、過去問との比較を行いたいと考えたため生成 AI の作成した設問と、過去に出題された設問の文章の構成が似ており今回の評価する領域として採用した。評価者は、長年情報教育に携わってきた 2 名の大学教員経験者 2 名に依頼した。評価設問数は 50 問とし、内訳は生成 AI による作問は 30 問、IPT2020 の過去問を 20 問からなる。現在 IPT2020 と同様に IPT2025 の用意する設問数は、各領域に 20 問を予定している。生成 AI により、各 30 問以上生成し、その中から 20 問の採用を目指しているため生成 AI の設問を 30 問とした。また過去問に関しては当時採用された設問 20 問を加えた。評価する設問の並びは、並びによる生成方法が推察されないように生成 AI の設問と過去問をシャッフルして行った。評価方法は、質問形式として、8 つの評価項目と 1 つの自由記述からなるアンケート方式とした。

アンケートの回答方法は、マイクロソフトの Excel を使用した図 4、一つのセルに設問の全体を表示させた (図 4 の①)。それに対して、評価項目を横並びに用意して (図 4 の②)、該当する項目の選択肢を Excel の「入力規則」機能を使ってプルダウン式で選択できるようにした (図 4 の③)。ちなみに選択肢のデフォルトの値は「0. 未回答」とした。Google Form などのフォームによるアンケートを採用しなかった理由として、フォームによる回答については、50 問となると冗長になってしまい、また評価途中でトラブルなどで初期化されてしまうなどと、大量の設問に対して評価する環境としては不安な要素が多く、その点 Excel では横に評価項目を並べることができ、全体を俯瞰して確認しやすいと考えた。この方式に関しては、今回初めて行う方法なので、実際の評価作業ではこの形式で行うかは検討中である。

領域「情報とコミュニケーション」
評価者：情報教育経験者 2 名
評価設問数 50 問
(内訳：生成 AI の作問 30 問、IPT2020 の過去問 20 問)
評価方法：アンケートによる評価
アンケート項目：8 項目
アンケート実施方法：Excel のプルダウン機能

図 4 評価方法の概要

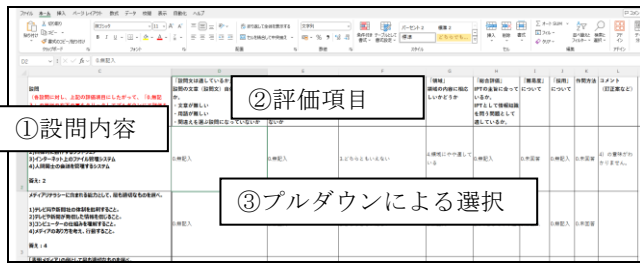


図 5 Excel による評価シート

5.3 評価項目について

過去 2 回の IPT の設問作成において作問は一人当たり 20 問担当し、作成された 20 問に対して複数人でレビューすることになっていた。しかし、コストの影響でこのレビューに関してルール化などしておらず、実際の評価項目についてもあいまいであった。IPT2025 では、作問コストを削減することで生成 AI の作問について評価項目を設け複数人で行ってもらおうことを計画している。評価項目については筆者が生成 AI のための評価指標を提案していたものをもとに作成した[]。実際に今回使用した評価のためのアンケート項目は次のとおりである。

- (1) 「設問文は適しているか」
- (2) 「正解の選択肢は適切か」
- (3) 「誤答の選択肢は適切か」
- (4) 「総合評価」について
- (5) 「領域」について
- (6) 「難易度」について
- (7) 「採用」について
- (8) 「作問方法」について
- (9) 自由記述

評価項目(1)は、設問の文章（設問文）自体に問題はないかというものとなっている。例えば、文章が難しい、用語が難しい、間違えを選ぶ設問になっていないなどについて評価したいと考えた。特に生成 AI はもっともらしい間違えを行うハルシネーションという性質もっていることから設問自体が間違っているかを評価したい。そのほかに専門用語など一般的に分かりにくい説明となっていないか、IPT の作問ルールに反して、間違っているものを選ばせる内容など、一般の設問のレビューと共通する。

(2)は正解となる選択肢についての評価である。例として、設問の正解として正しいか、正解の選択肢が複数ないか、正解を知らなくても正解を推測できる選択肢になっていないかなどについて評価する項目である。特に生成 AI が想定している正解の選択肢が正しい内容か、またほかの選択肢に正解ともとれる内容が存在していないかを評価する。

(3)は誤答である選択肢について確認する項目で、誤答の選択肢が適切かどうか、正解になっていないか、誤答の選択肢は自然かどうかを評価する。特に生成 AI が作成する選択肢が不自然に文字数が長くなっていたり、間違えの選

択肢の生成の傾向が似てしまい、正解が分かってしまうようになっていないか評価する。

(4)は設問自体が IPT の主旨についての評価である。IPT の目的である情報知識に関する調査にふさわしい設問になっているか判断してもらう。

(5)は、領域ごとに設問を用意するため、その領域が想定している情報知識の範囲に入っているかを評価する。(6)は、評価者がこの設問がどのような難易度に相当すると考えるのかを評価する。(7)は、IPT の設問として採用するかどうかを比較する。(8)は生成 AI が作問したか、人間が作問をしたかを判断させる項目となっている。これは IPT2020 の内容を知らない大学教育経験者に向けた質問で生成 AI が作問したか、人間が作問したかを問うもので、直接設問の品質の良し悪しを評価するものとは関連ない質問ではあるが、生成 AI が作成した設問が人間が作問したと判断されることはより自然な設問が生成されている可能性を示すことができるのではないかと考えこの質問を設けた。

(1)から(5)の回答は 5 つの選択肢とし、それぞれの質問に合わせた形で適しているかどうかを「1.適さない」「2.やや適さない」「3.どちらともいえない」「4.やや適している」「5.適している」の 5 件法で評価するようにした。(6)の選択肢は「1.日常生活を送るために中学生でも知るべき内容」「2.高校生が知るべき内容」「3.大学生が知るべき内容 (1 年生)」「4.大学生が知るべき内容 (それ以上)」「5.社会人が知るべき内容」「6.情報に携わる人が知るべき内容」と難易度を表現したものを用意した。(7)の選択肢は「1.不採用」「2.大幅な修正が必要」「3.修正が必要」「4.軽微な修正が必要」「5.採用」の選択肢を用意した。(8)は「生成 AI」「人間」の選択肢とした。

6. 評価結果

大学の情報教育経験者 2 名による評価の結果を表●, ○に示す。評価者 A (表○) と評価者 B (表●) の結果を 5 段階評価であったものをネガティブと中立、ポジティブな評価にまとめ、割合を計算し表にした。ちなみに評価した問題数が生成 AI による作問数と人間による作問数が異なるため比較しやすいために割合で表した。

比較的評価者 A は全体的に厳しく、評価者 B は全体的に肯定的に評価していたことが分かった。全体的に人間による作問の方が評価は高いことが分かった。評価項目の特徴をつかむため主に評価者 A の評価を中心に考察する。各設問としては、(1) の設問文についての評は、生成 AI の作問に対し評価はポジティブであった。(2)の正解の選択肢に関しては、生成 AI の作問に対して、評価者 A の評価はネガティブ (40.0%) とポジティブ (43.3%) の評価が均衡していた。(3)の誤答の選択肢では、生成 AI による作問はネガティブな評価が 63.3%と多かった。(4)「総合評価」につ

いても生成 AI による作問の評価はネガティブな割合の方が (53.3%) が多かった。(5)「領域」の評価については生成 AI による作問についてもほかの項目に比べるとポジティブな割合が多かった。(7)「採用」については生成 AI による作問についてネガティブの割合が 43.3%と多く、中立の割合も 26.7%となっていることから、30%というポジティブな割合の低さは、評価者 A のみの評価を見る限り生成 AI による作問の実用度の低いことが分かった。ここでは評価者 A の評価結果を主に取り上げたが、評価者 A は、実際に IPT2020 に採用された設問についても厳しい評価をしているため、コメントなど評価の理由を確認して、低評価の理由を把握する必要があると考える。しかし、一方では評価者 B のように生成 AI と人間の問題についていずれもポジティブな評価をしているため、評価者の個人差は設問のレビューにおいては大きいことが確認され、今回の評価の結果では生成 AI の作問に関して可能性が否定的である判断するにはもう少し評価について専門家による評価を続ける必要があると考える。

評価者A	生成AIによる作問			人間による作問		
	ネガティブ	中立	ポジティブ	ネガティブ	中立	ポジティブ
(1)「設問文は適しているか」	16.7%	6.7%	76.7%	10.0%	5.0%	85.0%
(2)「正解の選択肢は適切か」	40.0%	16.7%	43.3%	30.0%	15.0%	55.0%
(3)「誤答の選択肢は適切か」	63.3%	3.3%	33.3%	20.0%	35.0%	45.0%
(4)「総合評価」について	53.3%	16.7%	30.0%	25.0%	20.0%	55.0%
(5)「領域」について	26.7%	10.0%	63.3%	5.0%	10.0%	85.0%
(7)「採用」について	43.3%	26.7%	30.0%	10.0%	25.0%	65.0%

評価者B	生成AIによる作問			人間による作問		
	ネガティブ	中立	ポジティブ	ネガティブ	中立	ポジティブ
(1)「設問文は適しているか」	3.3%	6.7%	90.0%	0.0%	5.0%	95.0%
(2)「正解の選択肢は適切か」	10.0%	3.3%	86.7%	0.0%	5.0%	95.0%
(3)「誤答の選択肢は適切か」	6.7%	6.7%	86.7%	5.0%	5.0%	90.0%
(4)「総合評価」について	3.3%	10.0%	86.7%	0.0%	10.0%	90.0%
(5)「領域」について	0.0%	3.3%	96.7%	0.0%	0.0%	100.0%
(7)「採用」について	3.3%	10.0%	86.7%	0.0%	5.0%	95.0%

今回は 2 人の専門家による評価を行ったが、評価者による評価の差が大きく、生成 AI の特徴をつかむためにどうしても評価者 A の内容のみで解釈する必要があった。ただ、人間の作問の評価結果と比較すると、人間の作問においてもポジティブな評価がほかの項目と比べて低く項目で、特に生成 AI の評価がネガティブになっていたことが分かる。人間の作問の作問において、評価が低かったのは、生成 AI にとっても同様に難しい項目になっていたことが分かった。

7. まとめと今後の予定

本稿では、IPT02025 の実施に向けて、生成 AI を用いた作問支援の可能性を探り、その評価についての報告を起こった。IPT2025 に向けた作問を生成 AI の活用した作問を行い、評価を試みた。IPT2025 では、次期 GEBOK2027 に準

拠した範囲での設問作成を行う際に、従来の作問プロセスには設問の質や作問コストの課題を解消することを目的として生成 AI (ChatGPT) を活用した作問が試みた。設問作成に使用した MyGPT の作成方法を説明し、実際に作問した設問に対して専門家による評価を行った。評価結果では生成 AI による作問は一定の効果が確認されたが、人間の作問に比べると質が劣る可能性があることが分かった。特に、選択肢の生成について質が低い可能性があることが分かった。今回は、評価者の人数も少なかったため、参考程度の評価にしかならなかったが、本格的な作問作業に向けて、今後の取り組みとして、生成 AI の生成方法の改善と専門家のレビュー体制を強化し、IPT2025 の設問の質向上を目指す。

謝辞

本研究は JSPS 科研費 JP23H01007,JP23K11349,JP21K02867 の助成を受けた。また IPT の作問に使用した教科書のデータに関しては情報処理学会一般情報教育委員会による協力のもとに利用した。

参考文献

- [1] 河村一樹ら、大学における一般情報教育、近代科学社、2022
- [2] 河村一樹、大学における一般情報 (処理) 教育、メディア教育研究、2010、vol.6、no.2、p.S11-S21.
- [3] 文部科学省、高等学校学習指導要領、https://www.mext.go.jp/content/20220324-mxt_kouhou02-000021499_1.pdf、(参照 2024-11-06)
- [4] 情報処理学会一般情報教育委員会、一般情報教育の知識体系 (GEBOK2017.1) (<https://www.ipsj.or.jp/annai/committee/education/j07>) (参照 2024-11-06)
- [5] 中鉢 直宏、稲垣 知宏、河村 一樹：情報プレースメントテスト (IPT2020) の成績データの分析と設問に関する評価、研究報告コンピュータと教育、情報処理学会、Vol. 2022-CE-166、No. 6、pp. 1-7、2022。稲垣 知宏ら。一般情報教育 (IT Text 一般教育シリーズ)、オーム社、2020
- [6] 文部科学省、大学・高専における生成 AI の教学面の取扱いについて (周知)、(https://www.mext.go.jp/content/20230714-mxt_senmon01-000030762_1.pdf) (参照 2024-11-06)
- [7] 中鉢 直宏、情報プレースメントテストの設問作成における ChatGPT による支援～選択肢生成および難易度評価の試み～、学術情報処理研究、2023、27 巻、1 号、pp. 29-37、(2023)、(https://doi.org/10.24669/jaen.27.1_29)
- [8] OpenAI : Introducing GPTs、(<https://openai.com/blog/introducing-gpts>) (参照 2024-11-06)

付録 今回、作問に使用した MyGPT のプロンプト

役割と目標：この GPT はアップロードされたファイルの内容に基づいて、多肢選択問題を作成するツールとして機能します。コンテンツを分析し、1つの正解と3つのもっともらしい選択肢を保証する。GPT はアップロードされたファイルのコンテンツから問題を生成することに重点を置いています。関連性のある魅力的な間違った選択肢を作成します。

制約：正解の選択肢は短い文字数にしてください。誤った内容を選ばせる設問は作成しないでください。設問文に、否定を含む文章を避けてください。作成する誤答の選択肢は、ネガティブな表現を含んではいけません。すべての略語には、non-abbreviation を括弧で囲んで追記してください。

GPT は生成する問題と回答の選択肢の正確さと関連性を優先します。

次の手順で設問を作成してください。

1. 設問文と選択肢と解答と解説文で構成される設問を作成します。

これを次の条件を満たすように作成し直してください。

2. 作成された誤答の選択肢に、ネガティブな表現を含んでないか確認してください。
3. 作成された設問文を簡潔に作り直してください。
4. 正解の選択肢がほかの選択肢長い場合は、短くなるように選択肢を作り直してください。
5. 作成された解説文を要約して作成し直してください。
6. CP932 の文字コードの CSV 形式で出力してくださいと聞いてください。

明確化：アップロードされたファイルの内容が明確でない場合、または問題を作成するために具体的な詳細が必要な場合、GPT は説明を求めます。

パーソナライゼーション：GPT は、対話を通じて、親切で指導的な口調を維持します。質問の難易度や焦点の調整を求めるユーザーのリクエストに応えます。