

# 科目および課外活動内容からのスキル抽出と評価

尾崎 真大<sup>1)</sup>, 中條 麟太郎<sup>2),3),4)</sup>, 松尾 周汰<sup>4)</sup>  
ウン クアン イー<sup>1),4)</sup>, 大久保 文哉<sup>5)</sup>, 島田 敬士<sup>5)</sup>,

- 1) 九州大学 大学院システム情報科学府
  - 2) 東京大学 大学院学際情報学府
  - 3) 日本学術振興会, 4) 株式会社イマゴ
  - 5) 九州大学 大学院システム情報科学府
- ozaki.mahiro.493@s.kyushu-u.ac.jp

## Skill Extraction and Evaluation from Course and Extracurricular Activity Content

Mahiro Ozaki<sup>1)</sup>, Rintaro Chujo<sup>2),3),4)</sup>, Shuta Matsuo<sup>4)</sup>  
Kuan Yi Ng<sup>1),4)</sup>, Fumiya Okubo<sup>5)</sup>, Atsushi Shimada<sup>5)</sup>,

- 1) Graduate School of Information Science and Electrical Engineering, Kyushu University
- 2) Graduate School of Interdisciplinary Information Studies, The University of Tokyo
- 3) Japan Society for the Promotion of Science, 4) iMago, Inc.
- 5) Faculty of Information Science and Electrical Engineering, Kyushu University

### 概要

現状の高等教育機関において、学生が正課・課外活動を通じて獲得するスキルの可視化は限定的である。そこで本研究では、シラバス等の正課活動に関する文書に加え、大学内実務等の課外活動に関する文書を大規模言語モデル (LLM) に入力しスキル抽出を実行した。また、学生によるアノテーションを実施し、LLM による抽出精度を評価するとともに、正課・課外活動から得られるスキルの相互関係を明らかにした。その結果、LLM は高い精度と安定性で教育的文脈の文書からスキルを抽出できる有効な手法であることが確認された。また、正課活動と課外活動を組み合わせることで、互いに不足する一部のスキルを補い合う関係にあることが示唆された。一方で、文書の表面的記載から抽出されにくいスキルへの対応や、教育的文脈を想定した候補スキルリストの改良については今後の課題とした。

## 1 はじめに

近年の高等教育機関にとって、自らの責任で教育活動を評価・改善し、その質の保証を確立することが重要な課題となっている。そのため、文部科学省は「教学マネジメント指針」[1]を公表し、大学運営における5つの重要な取り組みの1つとして「学修成果・教育成果の把握・可視化」を掲げた。これにより、学生が「卒業認定・学位授与の方針 (ディプロマ・ポリシー: DP)」に定められた資質・能力を身に付けられていることを自ら実感し、説明できるようになることが期待されるとともに、大学運営においても教育課程の点検や改善への活用が求められている。

このような成果の可視化に向けた取り組みは、成績評価 (GPA) やルーブリック、授業評価調査、履修計画

調査の活用など多岐にわたるが、中でも特に学生が獲得すべき資質・能力、すなわち「スキル」に焦点を当てた研究や実践が活発に進められている [2]。例えば、全国の国立高等専門学校ではスキルベースの可視化が取り組まれており [3]、特に武市ら [4] はスキルシートと評価シートを用いた実験スキルの明確化と可視化に取り組んだ。また、Qenani ら [5] は“学生に自身のスキルを自覚させ言語化する機会を提供することは、大学の重要な責任である”とし、単なる知識に留まらない「スキル」を軸に高等教育を捉える重要性を主張した。

また、教学マネジメント指針は、「課外活動」の意義を学修者本位の教育を実現するという観点から積極的に捉え直し、各大学が適切にその支援等に取り組んでいくことを期待するものと位置付けている。愛媛大学が学生が身につけるべき能力として定めている「愛大

学生コンピテンシー」では、課外活動も含めた大学生生活全体の活動が包括的に考慮されており、特に、サークル活動やインターンシップ、海外留学などがいずれのスキルにも肯定的に関与すると論じられている [6]。また、近年の国内における大学生の新卒就職活動を踏まえても明らかのように、Yorke ら [7] はカリキュラムと雇用可能性 (employability) に関するスキルの密接な関連性を強調している。このように、当事者である学生を取り巻く環境を考慮すると、正課の授業に留まらず、課外活動を含めたあらゆる活動から獲得されるスキルを客観的に把握し、可視化することが不可欠であると言える。しかしながら、多くの高等教育機関において、こうした課外活動で培われるスキルも含めて学生が自身の成長を自覚できるような仕組みは十分に確立されておらず、課外活動とスキルの関連性に関する研究も限られている。

一方で、このような高等教育におけるスキルの可視化は、自然言語処理分野における「スキル抽出」とも関連している。スキル抽出とは、文書から職務遂行に必要な能力や個人の持つ能力を特定する情報抽出タスクであり、主に履歴書や求人情報といった人材・採用活動分野での応用を目的として研究が進められてきた。特に近年では、大規模言語モデル (LLM) の発展により、大規模なデータセットを用いずとも容易にスキル抽出タスクを実行できるようになるとともに、計算コストや速度の課題に対応した様々な手法が提案されている。例えば、LLM の要約能力とベクトル類似度検索を組み合わせた抽出手法 [8] や、LLM の文章生成能力を用いて作成した大規模なデータセットを用い、予測モデルをファインチューニングすることで、従来より高い精度を達成する方法論 [9, 10] が提案されている。

しかしながら、これまでのスキル抽出に関する研究は主に人材・採用活動分野に傾倒しており、教育の文脈、特に科目内容や課外活動内容からスキルを抽出した研究は限られている。人材・採用活動関連の文書と比較して、教育コンテンツの記述はより多様であり、学生が獲得できるスキルに関する表現も暗黙的かつ多岐にわたると考えられるため、既存のスキル抽出手法をそのまま適用できるとは限らない。また、教育分野でスキルの可視化に関する取り組みは存在するものの、教育文書からのスキル抽出に焦点を当て、その精度評価や学生目線での妥当性検証を実施した先行研究は、筆者の知る限り存在しない。

そこで本研究では、高等教育における学修成果をスキルとして可視化するための端緒として、LLM を活

用した正課活動および課外活動内容文書からの自動スキル抽出を試みる。具体的には、正課活動としての目標と科目内容を示した「学修目標」と「シラバス」に加えて、課外活動の一環であり大学内実務である「クエスト」(2.1.1 項で詳説) の募集要項を入力とし、LLM に各文書に紐づくスキルを抽出させる。そして、LLM によるスキル抽出の妥当性を評価するため、学生を対象に、文書に対応するスキルを選択するアノテーションを実施し、この結果を正解データとした精度検証を行う。さらに、このアノテーション結果を用いて、正課活動と課外活動からそれぞれ獲得できるスキルの偏りや関係性を明らかにする。

以上より、本研究の研究課題を以下の2つとする。

**RQ1** 正課・課外活動に関する文書からの自動スキル抽出において、LLM はどの程度の精度を示すか？

**RQ2** 学生の認識において、正課活動と課外活動から得られるスキルは互いにどのような関係にあるか？

これらの研究課題を明らかにすることで、「スキル」を軸として、正課・課外活動を通じた大学生の学修成果を捉える際の実現可能性と課題を明確にすることができる。最終的には、学生が自身の学修成果を客観的に把握できるシステムの設計や、教育機関によるカリキュラム改善の実現に資する知見の提供が期待される。

## 2 手法

本研究全体の流れは図1のとおりである。以下の項で使用するデータセット、LLM によるスキル抽出手法、学生によるアノテーション方法について詳説する。

### 2.1 データセット

#### 2.1.1 抽出対象の文書

本研究で取り扱う、科目内容および課外活動内容が記載された文書は以下の4種類152件である。なお、シラバス(基幹)は2024年度九州大学工学部I群(電気情報工学科)計算機工学コースのカリキュラムにおいて、学生が履修しうる科目を対象とし、「シラバス(専攻)」は同カリキュラムにおいて必修あるいは選択推奨とされる科目を対象とした。また、3節以降の説明のため、各文書を A1 から D4 までのカテゴリに分類した。

- **学修目標:** 計算機工学コースのカリキュラムに対して制定されたDPの一部であり、学生が身につけるべき知識や能力を具体的に示した指標である。カリキュラム上の全ての科目は1つ以上の学修目標と

紐づいている。

● A1 学修目標 (16 件)

- **クエスト:** 株式会社イマージョが運営する QuestGate (旧 Guild プラットフォーム)\*1 上で有償で九州大学生を対象に募集していた大学内実務 (クエスト) の募集要項。仕事内容やその目的、応募条件などの情報が含まれる。

● B1 アンケートなどの短期的なもの (26 件)

● B2 システム開発や運営メンバー募集などの長期的なもの (14 件)

- **シラバス (専攻):** 主に学部 2 年生以降を対象に開講される専攻教育科目の詳細な授業設計を示したもの。科目名、担当教員名、科目概要、授業計画、キーワードなどが含まれる。

● C1 “全分野共通基礎” に分類されるもの (19 件)

● C2 “数理情報科学”, “ソフトウェア工学” などの専門的な科目系統に分類されるもの (32 件)

- **シラバス (基幹):** 主に学部 1 年生を対象に開講される基幹教育科目の詳細な授業設計を示したもの。含まれる情報は上記と同様である。

● D1 英語や韓国語などの言語文化基礎科目 (12 件)

● D2 文系ディシプリン科目 (15 件)

● D3 理系ディシプリン科目 (13 件)

● D4 グループワークおよびその他の科目 (5 件)

### 2.1.2 抽出候補のスキルリスト

本研究では、教育分野に特化したスキル抽出とその評価を行うために、11 個のスキルから成る抽出候補スキルリストを独自に構築した。まず、Yorke ら [7] がカリキュラム設計や分析に組み込むべきスキルとして提示している合計 39 個をベースとして採用した。次に、解釈性や後述するアノテーションでの負担を考慮し、Tushar ら [11] が報告した主要なスキルリストを参考に、これらのスキルを 11 個の集合に集約した。各集合に対しては、同文献を参考に上位概念に当たるスキル名を定めた。最終的なスキルリストは以下の通りである。

1. **共同作業力:** “影響力”-自分の見解や行動方針を主張し、正当性を他者に納得させること。“チームワーク”-共通の課題に対して他者と建設的に協力できること。“政治的感性”-組織が実際にどのように機能しているかを理解し、それに依拠して行動すること。
2. **対人・コミュニケーション力:** “口頭発表力”-グループに対して明確かつ自信のある情報提示を行うこと。“説明力”-口頭および文書を通じて、明確に情報提示すること。“交渉力”-意見の相違のある問題について、相互に満足いく解決を図るための議論を行うこと。“感情的知性”-他者の感情とその影響に対する感受性を持つこと。“論証・正当化力”-自分の見解や行動方

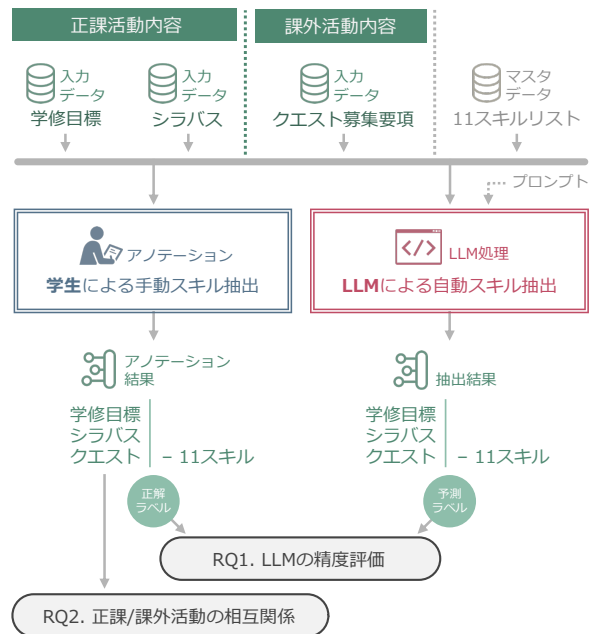


図 1: 本研究全体の流れ

針を主張したり、正当化したりすること。“書面での伝達力”-具体的に書かれた明確な報告書や手紙などを作成すること。“葛藤解決力”-自己内および他者との関係における葛藤を解決すること。“傾聴力”-重要なポイントを認識しながら集中して注意を払うこと。

3. **批判的思考:** “読解力”-重要なポイントを認識し記憶する能力。“批判的分析”-状況や課題の矛盾点を見つけ出す能力。
4. **グローバル・異文化対応力:** “グローバルな意識”-文化および経済の両面において、国外に関心を持つこと。“異文化間で働く能力”-国内外両方の文化圏において働く能力。“言語能力”-2 つ以上の言語を話す能力を持つこと。
5. **倫理観:** “倫理的感性”-就労の倫理的側面を理解し、それに依拠して行動すること。“道徳的行動力”-道徳的な規範を持ち、それに依拠して行動すること。“商業的意識”-ビジネスの課題や優先事項を理解して行動すること。
6. **自己管理能力:** “自己管理能力”-効率的かつ体系的に作業する能力。“優先順位付け”-重要度に応じてタスクのランク付けをする能力。“計画力”-達成可能な目標を設定し、行動を構築すること。“自立性”-他からの監督なしで行動する能力。“主体性”-促されることなく自発的に行動を起こす能力。“自己認識”-自分の強みと弱み、目標や価値観を理解していること。“反省性”-自分自身および他者のパフォーマンスについて評価的に振り返る傾向。
7. **問題解決力:** “問題解決”-解決策を見つけるために適切な方法を選択し使用すること。“複雑さへの対処”-あいまいな状況を処理する能力。“意思決定”-複数の選択肢の中から最良の選択をすること。“自信”-就労や生活の中で直面する課題に対処できるという自信。
8. **状況適応力:** “適応力”-変化する状況や新たな課題に対して積極的に対応する能力。“柔軟な自己理論”-自身の知能などは固定されたものではなく、発展できると信じられる気質。“ストレス耐性”-プレッシャーの下でも効果的に機能し続ける能力。
9. **創造性:** “創造性”-独創的または発明的な能力や、固定概念や既存の論理に捉われずに物事を多角的に考察する水平思考力。
10. **学習意欲:** “学習意欲”-就労や生活のニーズに応えるために継続的に学び続ける姿勢。
11. **情報処理能力:** “専門知識の応用”-高等教育プログラムで得た専門知識を活用すること。“コンピュータリテラシー”-様々なソフトウェアを使いこなす能力。“情報検索力”-様々な情報源にアクセスする能力。“数的能力”-正確に数字を扱う能力。

\*1 <https://imago.co.jp/guild>

## 2.2 LLM による自動スキル抽出

本項では、LLM を用いて、正課および課外活動に関する文書から 11 個の候補スキルを抽出する手法を説明する。LLM には OpenAI 社が API を介して提供する GPT シリーズのうち、フラッグシップモデルである *gpt-4.1-2025-04-14*、*gpt-5-2025-08-07*、*gpt-5-nano-2025-08-07* の 3 つを採用した。実装は Python の `openai` ライブラリを用い、レスポンス形式に JSON Schema を指定することで、出力を構造化形式に統一した。プロンプトは以下の通りであり、これにより 1 つの文書に対して 0 個以上のスキルが抽出される。また、生成結果のばらつきを考慮して、152 件の各文書に対して独立に 10 回のスキル抽出を実行し、各文書における各スキルの抽出割合 (0.0~1.0) を算出した。

# 科目および課外活動内容からのマルチラベルスキル抽出タスク

## 評価対象の文書

以下の文章は { 各種文書のタイトル } です。まずはこの文章を読んで、内容を理解してください。

{ 各種文書が Markdown 形式で入る }

## スキル抽出の実行

次に、学生がこの文書に書かれている内容を遂行することで、得られそうな「スキル」を以下の 11 個の中から 0 個以上選択してください。複数選択しても構いません。

- 共同作業力
- 対人・コミュニケーション力
- ……

スキルの詳細な説明は以下の通りです。

- \*\*共同作業力\*\*: チームワーク - 共通の課題に対して他者と建設的に協力できること。影響力 - 自分の見解や行動方針を主張し、正当性を他者に……
- \*\*対人・コミュニケーション力\*\*: 説明力 - 口頭および文書を通じて、明確に情報提示すること。口頭発表力 - グループに対して明確かつ自身の……

## 注意事項

- 詳細なスキル名ではなく、それらを含めたスキル名を選択してください。
- 上記 11 個以外のスキルを出力しないでください。
- 同じスキルを重複して選択しないでください。
- `response_format` の `skill_schema` に従って json 形式で出力してください。

## 2.3 学生によるアノテーション

本研究におけるスキル抽出タスクの有効性を検証するため、複数人の学生を対象に、文書に対応するスキルを選択するアノテーションを実施した。すなわち、2.2 項で述べた LLM によるスキル抽出を、人力で行うことに相当する。アノテーションには、九州大学工学部 I 群 (電気情報工学科) 計算機工学コースに所属する学部 3, 4 年生、および同コースを卒業済みの大学院生の合計 48 名が参加した。参加者の負担を軽減しつつ作業の精度を確保するため、図 2 に示すアノテーションツールを独自に実装した。まず、参加者はツール内のチュートリアルを経て、タスクの内容と 11 個の候補スキルの定義を厳密に確認した。その後、抽出対象の文書を読み、習得できると考えられるスキルを



図 2: 作成したアノテーションツール

複数選択するタスクを、全 152 個の文書に対して実施した。最終的に収集された 48 名分の結果より、各文書における各スキルの抽出割合 (0.00~1.00) を算出した。

## 3 結果

### 3.1 LLM によるスキル抽出の精度検証 (RQ1)

LLM および学生によるスキル抽出結果の整合性を確認するため、本タスクを LLM によるマルチラベル分類タスクと見なして評価した。具体的には、48 名の学生がアノテーションした結果のうち、抽出割合が過半数を超えた (閾値  $\theta = 0.5$ ) スキルを正解ラベル集合とした。また、LLM の出力については、10 回の実行における抽出割合を予測データとして扱った。そして、各 GPT モデルおよびスキルごとの評価指標として、ROC 曲線の AUC と、LLM の抽出割合の閾値  $\hat{\theta} = 0.5$  における F1 スコアを用いた。結果を表 1 に示す。support (正解データにおける正例の数) のばらつきを考慮し、Macro 平均に加えて、Micro 平均と support で重み付けした Weighted 平均を記載している。

#### 3.1.1 LLM ごとの抽出精度

各 GPT モデルの Weighted 平均を比較すると、F1 スコアでは GPT-4.1 が最も高い精度 (0.77) を示し、AUC についてはいずれのモデルも 0.8 を超える高い精度を達成した。このことから、概して高い精度でスキル抽出が可能なことを示しており、特に、GPT-4.1 モデルがスキル抽出に適していることを示している。また、GPT-5 シリーズにおいて AUC と F1 スコアに差が見られることから、予測結果の抽出割合の閾値  $\hat{\theta}$ 、あるいは正解データの正例とする閾値  $\theta$  を調整することで、より高い F1 スコアを達成できる可能性を示唆している。

表 1: LLM によるスキル抽出の精度評価

候補スキル	support	GPT-4.1		GPT-5		GPT-5-nano	
		AUC ↑ (ROC)	F1 ↑ ( $\hat{\theta} = 0.5$ )	AUC ↑ (ROC)	F1 ↑ ( $\hat{\theta} = 0.5$ )	AUC ↑ (ROC)	F1 ↑ ( $\hat{\theta} = 0.5$ )
共同作業力	26	0.90	0.79	0.91	0.73	0.93	0.67
対人・コミュ	51	0.95	0.90	0.93	0.83	0.94	0.79
批判的思考	34	0.57	0.38	0.68	0.41	0.68	0.40
グローバル	20	0.98	0.80	0.98	0.85	0.89	0.76
倫理観	5	0.89	0.67	0.87	0.47	0.87	0.38
自己管理能力	66	0.70	0.64	0.68	0.67	0.65	0.66
問題解決力	87	0.90	0.88	0.91	0.89	0.85	0.83
状況適応力	27	0.61	0.38	0.65	0.39	0.63	0.37
創造性	16	0.89	0.73	0.94	0.67	0.93	0.71
学習意欲	100	0.89	0.90	0.73	0.45	0.88	0.85
情報処理能力	81	0.82	0.82	0.82	0.82	0.70	0.72
Macro 平均		0.83	0.72	0.83	0.65	0.81	0.65
Micro 平均		0.86	0.75	0.84	0.68	0.87	0.70
Weighted 平均		0.83	0.77	0.81	0.68	0.80	0.72

### 3.1.2 スキルごとの抽出精度

スキルごとの結果に注目すると、「対人・コミュニケーション力」「グローバル・異文化対応力」「問題解決力」は、いずれのモデルにおいても高いスコアを示す傾向が見られた。一方で「批判的思考」「状況適応力」「自己管理能力」は、いずれのモデルにおいてもスコアが低い傾向が見られた。それ以外のスキルについては、モデル間でばらつきがあるものの、最も高い精度を達成した GPT-4.1 においては AUC が 0.8 を超える高精度であった。また、「倫理観」と「創造性」は、AUC に対して  $\hat{\theta} = 0.5$  における F1 スコアが極端に低くなった。これは、2つのスキルにおける support 数が全体の 152 件に対して極端に少ないことに起因する可能性があり、この点には留意する必要がある。

### 3.2 正課/課外活動から得られるスキルの関係 (RQ2)

正課活動と課外活動から得られるスキルの相互関係を確認するため、学生のアノテーション結果を文書カテゴリ別に分析した。具体的には、正課活動に該当する「基幹科目」「専攻科目」と、課外活動に該当する「短期クエスト」「長期クエスト」の4つの文書カテゴリにおけるスキル抽出割合をスキルごとに平均し、その結果を図3のレーダーチャートとして可視化した。

図が示すように、いくつかのスキルにおいては正課・課外活動のいずれかで相補的に獲得されていることが確認された。例えば、「対人・コミュニケーション力」「共同作業力」「創造性」は正課活動では割合が低く、獲得しにくいと評価されたが、課外活動によって補完されている。一方、「学習意欲」や「グローバル・異文

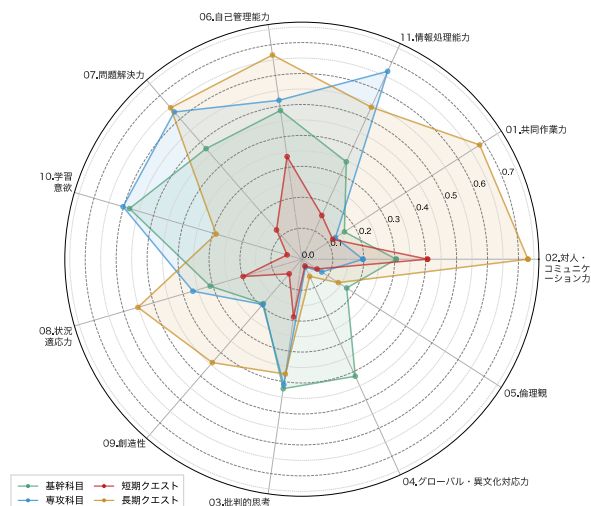
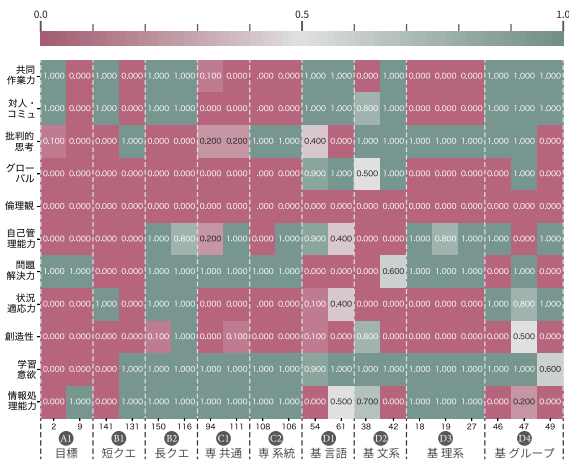


図 3: 正課・課外活動の文書ごとのスキル抽出割合

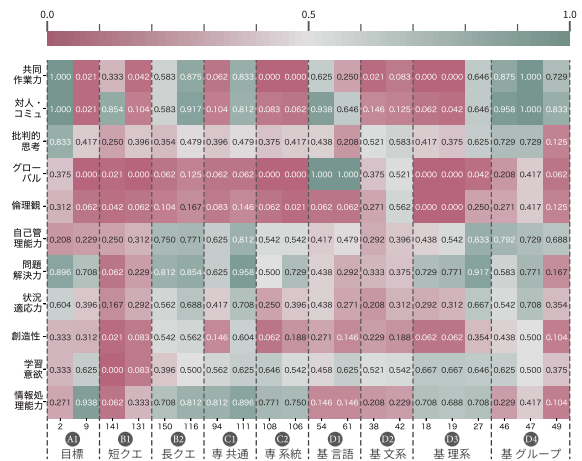
化対応力」は課外活動だけでは不十分であり、正課活動がその不足を補っている。ただし、「倫理観」については、正課・課外活動のいずれにおいても抽出率が低いままであった。以上より、全てのスキルを完全に網羅するとは限らないものの、正課活動と課外活動を組み合わせることで互いに不足する部分を補い合う関係にあると考えられる。

## 4 考察

本節では、LLM による自動スキル抽出結果と学生によるアノテーション結果を比較・考察する。GPT-4.1 モデルによる抽出結果の一部を図 4a に、学生によるアノテーション結果の一部を図 4b に示す。各ヒートマップは、11 個のスキルを縦軸に、各カテゴリから抜



(a) GPT-4.1 によるスキルごとの抽出割合 (抜粋)



(b) 学生によるスキルごとの抽出割合 (抜粋)

図 4: 各スキル抽出の結果

2: 目標-協働的な問題解決 | 9: 目標-工学的解析と実装 | 141: 車利用に関する集団インタビュー募集 | 131: 学習教材内の単語ラベリング作業募集 | 150: チャットボット開発者募集 | 116: エンジニア募集 | 94: データベース | 111: 専門実験 | 108: 分散システム | 106: 数理計画法 | 54: 英語 (speaking) | 61: スペイン語 | 38: 社会学入門 | 42: 政治学入門 | 18: 線形代数学 | 19: 力学基礎 | 27: 自然科学総合実験 | 46: 私にとっての学び・発表 | 47: 課題協

粋した文書 ID を横軸に配置し、過半数である 0.5 を境に色分けした抽出割合を示したものである。

#### 4.1 LLM の抽出精度に関する定性的考察 (RQ1)

3.1 項で述べたように、本タスクを LLM によるマルチラベル分類タスクと見なして評価した結果、特に GPT-4.1 は全体平均として高い精度を達成した。一方で、一部のスキルにおいてはスコアが低い傾向が見られた。そこで、図 5 に示すスキルごとの陽性・陰性の分布を用いて、定性的に分析した。

まず、高い精度を得た「対人・コミュニケーション力」「グローバル・異文化対応力」「問題解決力」については、その要素を示すキーワードや文脈が、文書中に明確に記述されていることが推察される。例えば、B2 長期クエストや D4 グループワーク系基幹科目では、“現状の課題に向き合う”や“学生及び教員との対話”などの表現が見られた。このような明確な判断基準によって、LLM と学生の双方がこれらのスキルを抽出しやすくなり、真陽性 (TP) となる傾向にあったと考えられる。

一方で、低スコアに留まった「批判的思考」と「状況適応力」については、これらのスキルが文書の表面的な記述から判断しにくく、より深い文脈理解と推論が必要となるためと推察される。特に「批判的思考」は、C2 専門科目において偽陽性 (FP) として誤分類される例が顕著であった。この原因として、シラバス上では批判的議論の余地があるように記述されているにもかかわらず、実際の授業では一方向的な知識伝達に終始しているケースが考えられる。また「状況適応力」について

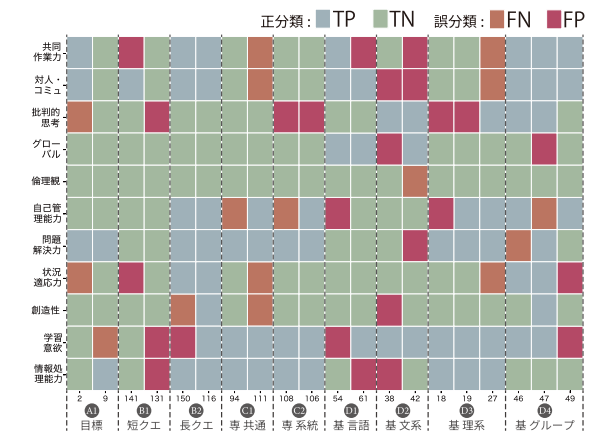


図 5: スキルごとの陽性・陰性の分布 (抜粋)

文書 ID については図 4 と同様

も、(141) のインタビューに類似する短期クエストにおいて、同様に偽陽性 (FP) として誤分類された例が多かった。これは、実態が単純作業であるにもかかわらず、募集要項内の“〇〇をいくつかの観点で評価”のような記載によって、LLM が過度に状況適応力が必要だと判断したためと考えられる。

総じて、LLM によるスキル抽出の精度は、シラバスや募集要項などの文書上の記述内容と、学生が実体験を通じて認識する授業形態や活動実態との間の差異に依存すると考察される。特に、文書が科目や実務の概念的内容に深く言及する一方で、学生の具体的な経験目線での情報が不十分である場合、LLM は表面的な情報に依拠せざるを得ず、スキルの過剰あるいは過少な抽出を招く傾向が示唆される。

## 4.2 正課/課外活動の関係に関する定性的考察 (RQ2)

図 4b の学生による抽出割合の傾向を確認し、3.2 項で述べた正課・課外活動から得られるスキルの相互関係について考察する。まず、課外活動を中心に抽出された「対人・コミュニケーション力」「共同作業力」については、特に B2 長期クエストにおいて高い抽出割合を示すことが確認された。これは、長期クエストのほとんどが、複数人が密接に関わり合う活動であることを反映している。一方、(111), (27) の実験系科目や D4 グループワーク系基幹科目を除いて、単独遂行型の授業が大半を占める正課活動では、これらのスキルの抽出割合は低い傾向にあり、課外活動が正課活動を補完する関係になったと考えられる。

続いて、正課活動を中心に抽出された「学習意欲」については文書カテゴリに依らず多く抽出され、「グローバル・異文化対応力」については D1 言語文化基礎科目で顕著に抽出された。これらのスキルはクエストの募集要項においてほとんど言及されておらず、実務よりも学術的な場で獲得する傾向が強いため、正課活動が課外活動を補完する関係になったと考えられる。

一方で、「倫理観」については、表 1 の support 数からも明らかなように、全体を通して抽出割合が低い結果となった。このことから、今回扱った 152 件の文書において「倫理観」を抽出することは困難であり、そもそも倫理観を抽出対象のスキルに含めることが適切でない可能性が示唆される。

## 4.3 スキル抽出の安定性

GPT モデルおよび学生のスキル抽出における安定性の傾向を確認するため、スキルごとの標準偏差を表 2 にまとめた。まず、各モデルの列を比較すると、モデル間で抽出結果の安定性が異なることが確認された。具体的には、GPT-4.1 は全体的に標準偏差が小さい一方、GPT-5 と GPT-5-nano は全体的に標準偏差が大きい傾向を示した。この結果は、各モデルの設計思想の違いに起因すると考えられる。OpenAI 社のリリースノート<sup>\*2\*</sup><sup>\*3</sup>によると、GPT-4.1 は長文テキストに特化し、指示追従性を高めるよう設計されているため、より一貫性のある出力を生成したと推察される。一方で、GPT-5 シリーズは、より複雑な推論を含む多様な用途に向けて設計されており、単一の明確な答えよりも多様な解釈を許容する傾向が強まることが示唆される。

表 2: 各スキル抽出結果における標準偏差

候補スキル	GPT-4.1	GPT-5	GPT-5-nano	student
共同作業力	0.01	0.04	0.15	0.22
対人・コミュ	0.04	0.05	0.17	0.33
批判的思考	0.07	0.12	0.23	0.45
グローバル	0.02	0.01	0.04	0.15
倫理観	0.01	0.02	0.08	0.24
自己管理能力	0.10	0.11	0.19	0.47
問題解決力	0.04	0.06	0.15	0.42
状況適応力	0.08	0.08	0.25	0.45
創造性	0.04	0.06	0.14	0.35
学習意欲	0.04	0.19	0.15	0.43
情報処理能力	0.06	0.07	0.14	0.41
平均	0.05	0.07	0.15	0.41

次に、学生によるスキル抽出結果の標準偏差に注目すると、LLM と比較して顕著に大きく、アノテーションの一貫性に欠ける傾向が示された。この原因として、「批判的思考」「問題解決力」「状況適応力」といった抽象度の高いスキルや、「自己管理能力」「学習意欲」「情報処理能力」といった、解釈の個人への依存度が高いスキルが挙げられる。例えば、ある科目において「情報処理能力」を獲得可能と判断するか否かは、アノテーションを行う学生が既に有する能力や、授業に対する意識によって異なる。このように、科目内容とスキルとの関連付けは、アノテーションの主体である各学生の背景によって大きく影響を受けることが示唆された。

## 5 まとめ

本研究は、高等教育機関において、学生が正課活動と課外活動の両者を通じて獲得するスキルの可視化が不足しているという課題に対し、LLM を活用した自動スキル抽出の有効性を検証しつつ、学生の認識における、正課活動と課外活動から得られるスキルの相互関係を明らかにした。具体的には、正課活動を記述したシラバスや学修目標、および課外活動を記述したクエストに関する文書を対象に、LLM によるスキル抽出と学生によるアノテーションを実施し、その結果を比較・評価した。

その結果、学生によるアノテーションを正解データとして評価したところ、LLM によるスキル抽出は全体平均として高い精度を示し、教育的文脈の文書からスキルを抽出する手法として有効であることが確認された。しかし、スキルごとの精度に着目すると「批判的思考」などの一部のスキルはスコアが低い傾向にあ

\*2 <https://openai.com/index/gpt-4-1>

\*3 <https://openai.com/index/introducing-gpt-5>

り、これは文書上の記述内容と、学生が実体験を通じて認識する情報との間の差異によって生じると考えられる。文書の表面的記述から判断することが困難なものについては、LLMにより深い推論を実行させたり、他のデータソースを探索したりする方法を検討する必要がある。また、正課活動と課外活動の文書カテゴリごとのスキル抽出割合を分析した結果、「共同作業力」などの多人数とのインタラクションに関するスキルは正課活動だけでは不十分であり、課外活動がその不足を補完する関係にあることが示唆された。このことから、正課活動に加えて課外活動も考慮に入れた包括的な観点から、学修者目線でスキルを捉えることの重要性が明らかになった。

一方で、学生によるアノテーション結果において、一部の抽象度の高いスキルでばらつきが大きく、正例が少ないスキルも存在するという課題も残された。これは、先行研究からトップダウン的に構築されたスキルリストが、教育コンテンツの多様な記述と適合しなかった可能性が考えられる。今後は、シラバスなどの教育的文書からの抽出を想定して、スキルリストをボトムアップ的に改良したり、シラバスでは測りきれないスキルについては他のデータソースの活用可能性を検討したりする。また、課外活動として扱ったデータが限定的であった点も課題である。将来的には、学生が参加したインターンシップやサークル活動などのより多様なデータを収集し、学修成果の包括的な可視化に取り組むことが期待される。

**謝辞** 本研究は、九州大学と株式会社イマージとの共同研究の成果の一部である。また、JST CREST JPMJCR22D1, ならびに JSPS 科研費 JP22H00551, JP22K19835 の支援を受けて実施された。最後に、リサーチ・アシスタントとしてご尽力いただいた九州大学の宗村隆世氏、橋谷田航太氏、藤原晃来氏に深く感謝申し上げる。

## 参考文献

- [1] 文部科学省 中央教育審議会大学分科会：教学マネジメント指針, (オンライン), [https://www.mext.go.jp/content/20200206-mxt\\_daigakuc03-00004749\\_001r.pdf](https://www.mext.go.jp/content/20200206-mxt_daigakuc03-00004749_001r.pdf) (2020). accessed 2025-09-26.
- [2] Shoaib, M., Abdullah, F. and Ali, N.: A research visualization of academic learning skills among students in higher education institutions: a bibliometric evidence from 1981 to 2020, *Library Philosophy and Practice*, Vol. 5579, pp. 1–34 (2021).
- [3] 独立行政法人 国立高等専門学校機構：モデル

コアカリキュラム, (オンライン), [https://www.kosen-k.go.jp/nationwide/main\\_super\\_kosen](https://www.kosen-k.go.jp/nationwide/main_super_kosen) (2023). accessed 2025-09-26.

- [4] 武市 義弘, 與那嶺 尚弘, 鈴木 真ノ介, 山田 親稔, 野口健太郎, 矢島 邦昭, 佐藤 淳: 工学実験・実習のスキルを可視化する手法, *電気学会論文誌D (産業応用部門誌)*, Vol. 136, No. 10, pp. 680–685 (2016).
- [5] Qenani, E., MacDougall, N. and Sexton, C.: An empirical study of self-perceived employability: Improving the prospects for student employment success in an uncertain environment, *Active Learning in Higher Education*, Vol. 15, No. 3, pp. 199–213 (2014).
- [6] 松本 長彦: 愛媛大学学生として期待される能力～愛大学生コンピテンシー～を解説する (試論), *大学教育実践ジャーナル*, Vol. 11, pp. 1–10 (2013).
- [7] Yorke, M. and Knight, P.: *Embedding employability into the curriculum*, Learning and employability, Higher Education Academy (2006).
- [8] Li, N., Kang, B. and Bie, T. D.: SkillGPT: a RESTful API service for skill extraction and standardization using a Large Language Model, *arXiv preprint arXiv:2304.11060* (2023).
- [9] Decorte, J. J., Van Haute, J., Deleu, J., Develder, C. and Demeester, T.: Design of negative sampling strategies for distantly supervised skill extraction, *Proceedings of the 2nd Workshop on Recommender Systems for Human Resources (RecSys-in-HR 2022)*, Vol. 3218, CEUR, p. 7 (2022).
- [10] Decorte, J. J., Verlinden, S., Van Haute, J., Deleu, J., Develder, C. and Demeester, T.: Extreme multi-label skill extraction training using large language models, *AI4HR & PES 2023 : International Workshop on AI for Human Resources and Public Employment Services, Proceedings*, pp. 1–10 (2023).
- [11] Tushar, H. and Sooraksa, N.: Global employability skills in the 21st century workplace: A semi-systematic literature review, *Heliyon*, Vol. 9, No. 11 (2023).