

生成 AI で大学数学はどこまでできるか？

線形代数を中心とした現状分析と教育への示唆

大阪公立大学・国際基幹教育機構* 吉富 賢太郎

Kentaro Yoshitomi

Faculty of Liberal Arts, Sciences, and Global Education

Osaka Metropolitan University

§1. 序論

2022 年末の ChatGPT の登場以降、大規模言語モデル (LLM) をはじめとする生成 AI の技術的進化は目覚ましく、社会全般のみならず高等教育機関における学習・評価環境にも大きな影響を与えている。生成 AI はあらゆる課題に対応可能であるかのような印象を与えがちであるが、生成 AI の本質的な特性には、なお解明されていない点が多い。文脈に応じた適切な解説文の生成や単純な知識補完においては大学生向けの学習支援ツールとして十分な実用性を有する一方で、論理構造の厳密な検証や、教員の出題意図に合致した論法による解説の構築という点では、依然としてハルシネーションのような不正確さや強引に計算して正解を得るような非効率的な解法が散見される。すなわち、少なくとも本稿で扱う範囲では、現在の生成 AI は「論理的な思考を自律的に行う知能」というよりは、提供された事実や学習データに基づく「超高度な検索・構成機能」の域を出ていないという側面が強い。

高等教育における数学教育、とりわけ抽象的な概念理解と具体的な計算スキルが交錯する線形代数教育において、この生成 AI の「能力の境界線」を正確に把握することは急務である。教員が意図する教授学的視点からの解説生成にはなお距離があるものの、学生が主体的に生成 AI と対話する環境をデザインできれば、メタ認知を促進する強力な形成的

* 〒536-8525 大阪市城東区森之宮二丁目 1 番 132 号

評価 ([1]) のツールとなり得る。

本稿では、生成 AI に問題を与えた場合に、適切に正解を導出できる問題、正解に至っているが解答過程が教育的に不適切な問題、不正解となる問題の境界線を具体例を用いて検証する。また、その結果を踏まえた AI 共生時代におけるオンラインテスト（主に数式自動採点システムの一つである STACK^{*1}）を利用した問題設計、および証明問題の学習支援における活用可能性についても論じる。以下では、単に AI と書いた場合も生成 AI を指すものとする。また、オンラインテストやオンライン演習のための問題をオンライン問題と記すこととする。

なお、本稿記載の内容は可能な限り正確を期したが、AI は日々進化しており、随時検証が必要であることを申し添える。

§2. 教育における AI の有用性と課題

ここでは、まず教育における AI の有用性と課題について検証すべきポイントを整理する。始めに先行事例を紹介する。

AI 活用の先行事例

神戸大学の長坂耕作氏は、反転授業の予習教材として、Moodle の Lesson 形式（自動生成された階層型多肢選択問題）を利用しており ([4])、フィードバックに AI へのプロンプトを出力して「理解が困難な場合には、このプロンプトを用いて説明を参照する」よう案内している。間違った説明が出ないようにプロンプトをあらかじめ調整している。

筆者の場合（2025 年後期）は、毎回のアンケートに「不明点がある場合には AI に質問し、内容を要約する」という設問を設置し、また、学生の要望もあって、全員の「まとめ」を閲覧できるようにページとして追加するなどしている ([5])。

AI の活用と弊害

AI の利用を全面的に禁止することは現実的ではない。一方、逆に教員が AI の利用を促進しようとしても、学生の多くはまだ「AI を信頼しない」状況であることがいくつかの研究で報告されている。実際、筆者が実施した授業アンケートでも「なぜ AI を使わないのか」という質問を数年にわたり実施した結果、同様のコメントが得られている。理由と

^{*1} <https://stack-assessment.org/>

しては、主に「AI は嘘をつくから信用できない」という意見が多数を占めるが、AI の解説を取得・閲覧したときの不満点として、「未習の内容を用いた説明のため理解できない」という意見も多い。Local LLM や高度な検索拡張生成 (RAG)、あるいは高度なファインチューニングは不要としても、履修内容や履修順序を予め適切に組み込んだプロンプトの設計が必要と考えられる。

形成的評価においては、学習者の学習に対するフィードバック ([2], [3]) が重要な要素である。学習者が理解に困難を抱える内容について、AI が詳細な解説を提示することができれば、AI はフィードバックを提供するツールとして学習支援に資する存在となる。教員にとっても、前述の STACK のような数式自動採点システムを利用している場合に、フィードバック作成の負担を軽減できる。あるいは、教員のフィードバックが学生に適切に理解されない場合には、AI に補足的な解説を生成させることで、オフィスアワーを補完する機能を果たすことも期待される。

以上のように、数学教育におけるフィードバックならびに解説支援の観点から、AI は教員と学生の双方にとって有用なツールとなり得る。

一方、期末試験のように AI の参照が認められない評価場面^{*2}は別として、自宅等で小テストとして課すオンラインテストを実施する場合はどうであろうか。

一般に、毎回の授業で小テストを実施することは学生の理解度を確認する上で重要であり、多くの大学教員が実施していると考えられるが、採点には相当量の労力を要する。ティーチングアシスタント (TA) に委託する例もあるが、時間コストが高い、修正作業が必要となる、など問題もあり、一般に効率的ではない。また、学習者にとって重要なのはフィードバックであるが、紙のテストの場合、答案毎にフィードバックを記述する負担は大きい上、学生が確認できるのは一週間後となる。さらに、一週間経過すると前週の学習内容に関する記憶内容の希薄化によりフィードバックの教育効果の低減が危惧される^{*3}。

これらの問題を解消するためには、即時フィードバックが可能なオンライン自動採点システムが重要である。筆者の所属環境では STACK を利用できるため、過去 10 年以上にわたり STACK によるオンライン問題を活用している。

逆に、このようなオンライン問題を総括的評価に用いる場合には、学生が AI を利用して解答する可能性があり、評価の信頼性を損なう要因となり得る。AI の性能は継続的に向上しており、AI の主たる情報源であるウェブ検索では解法の入手が困難であるような、

^{*2} ただし、AI グラス等の機器を用いた不正利用は今後注意が必要である。

^{*3} 最近では、PDF2submission (https://moodle.org/plugins/block_pdf2submission) のような答案返却システム等を用いれば、次回授業までに返却も可能だが、実際には学生が返却内容を確認する割合は高くないと考えられる。

AI 利用に対抗し得ると期待される問題の恒常的な開発にも、当然限界がある。

以上のように、AI は教育を支援する一方で評価上の課題も生じさせる。したがって、AI の能力範囲の把握が、AI の活用方法および対策を適切に検討する上で重要である。

以下本稿では、上述の観点から具体的な事例に基づき検証を行う。

§3. AI の能力範囲

具体的な検証に入る前に、本論の前提を整理する。想定するのはオンライン問題を活用する授業モデルである。AI 共生時代のオンライン問題のポイントは次の 2 点である。

1. 生成 AI によるフィードバック（解説）は適切か？
2. 生成 AI を利用可能な状況でオンラインテストを機能させるには？

ここでは、補助的に「AI に学生の誤答パターンは生成できるか？」という観点も追加するものとする。これは、オンライン問題をデザインする場合に AI により生成された誤答パターンが参考になるのか、を問うものである。以下、検証した結果の総括的な特徴を述べた後、次節で具体的な検証事例を紹介する。検証した事例は、主に STACK によるオンライン問題を対象とし、AI がどこまで解けるか、教員の意図通りの解説をするか、という観点からオンラインテストおよびオンライン演習における有効性を検証する。また、証明問題^{*4}に関して、学生の誤りパターンを生成できるか、を検証する。

AI に直接解答を求めた場合の解答・解説の特徴と影響

AI に計算問題の解答を直接求めた場合、概して以下の特徴および影響が認められる：

- 生成 AI 誕生当時は計算に弱点が認められたが、現在は Python を呼び出す（コードも明示する）ことにより正確に計算する。したがって、オンラインテストにおいて単純な計算問題のみを用いることは適切ではない。
- 計算問題はほぼ正答する一方、解説が出題意図に沿わない場合が多い。フィードバックとしての利用時には十分注意が必要である。
- 論証については、簡単なもの（基本的にはインターネット上にあるようなもの）は解説に大きな問題は見られない。例として、学生が困難を感じやすい一次独立性に

^{*4} STACK では、Parsons 形式という解答形式があり、証明問題を一種の順序並び替え問題として提示する機能がある。

関する証明にも概ね対応できる。

- ハルシネーションや不正確さは依然として存在する（今後も皆無にはならないと見込まれる）。誤りを指摘すると修正することがある一方、指摘を重ねた場合には自己正当化に相当する応答を示すことがある。
- AI システム（Gemini, ChatGPT, Claude, Copilot 等）、モデル、利用時期、有料版・無料版の別によって正答率には差が認められる。ただし、正答した場合の説明内容には、一般に顕著な差は認められない。

線形代数において不適切な解答・解説をする問題の特徴は筆者の検証した範囲では主に以下の通りである：

- 行列式に関する問題が多い（多重線形性の理解が不十分である等）。
- 整数条件の処理に弱点がある（整数成分を要求しているにもかかわらず小数を提示する等）。
- 理論的には簡単な問題に対しても、反復計算によって解を求める傾向がある（最終的な解答は正しい場合がある）。
- 「ある条件を満たす例を挙げよ」といった構成問題では誤答が多い。ただし、具体例ではなく一般的な構成法を問うと正答する場合もある。

§4. 例

本節では具体的な検証例を示す。使用した AI の種類や無料版か有料版かについては可能な限り正確に記載したが、バージョン（使用モデル）を特定できなかった場合があり、その場合はその時点での標準モデルで記載した。また、冒頭でも述べたように、本稿で示す AI の応答は、検証した時点の各モデルによるものであり、同一プロンプトであっても今後異なる応答を返す可能性がある点に留意されたい。

AI が正解する例・不完全解答する例 1

例 1 (多肢選択問題). 2 次正方行列 $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ に対して、次の値を対応させる関数のうち、列に関して多重線形性を持っているものをすべて選びなさい。

(1) $ad - 2bc$

- (2) $-a - 2b - c + d$
- (3) $2ac - 2bd$
- (4) $-3ad - bc$
- (5) $a^2 + b^2 + c^2 + d^2$
- (6) $-ab - 2cd$

この問題を ChatGPT に解答させると、以下のように、(2) $-a - 2b - c + d$ を正解として選択する。理由を尋ねると、「単なる一次式であってもよいと判断した」旨の応答を示す。これはおそらく、各変数への線形性と行列の列に関する線形性の違いが曖昧になっているためであろう。2026 年 2 月 22 日における別の AI での検証では、Gemini (Gemini 3-Flash/無料プラン), ChatGPT (GPT-5/Pro プラン), Microsoft 365 Copilot (GPT-5 ベース (当時) / エンタープライズ データ保護) は正解したが、Claude (Sonnet 4.6/無料プラン) は不正解 (同じ間違い), Claude (Opus 4.6/Pro プラン) は不正解 (異なる間違い) であった。このように、バージョンや時期 (おそらく、学習したデータの有無等) によって結果が変化することを示す顕著な例である。なお、GPT-5 Instant は当時の ChatGPT の標準モデルであり、GPT-5 Instant 以外のモデルを使用した可能性は排除できない。

ChatGPT (GPT-5 Instant, Pro Plan, 2025 年 9 月) の解答

- (1) $ad - 2bc$: a, d (第 1・第 2 列の成分) に 1 次, b, c にも 1 次. 各列に対して線形. **多重線形**
- (2) $-a - 2b - c + d$: 各変数が 1 次, 項の積がない. もちろん各列に線形. **多重線形**
- (3) $2ac - 2bd$: a, c は第 1 列, b, d は第 2 列. 第 1 列に対しては非線形 (ac が同じ列内の積). **多重線形でない**
- (4) $-3ad - bc$: a, d は異なる列, b, c も異なる列. どちらの項も各列をまたいだ 1 次の積. **多重線形**
- (5) $a^2 + b^2 + c^2 + d^2$: 各列に対して 2 次. **多重線形でない**
- (6) $-ab - 2cd$: a, b は異なる列, c, d も異なる列. 各列に対して線形. **多重線形**

正解: $ad - 2bc$, $-a - 2b - c + d$, $-3ad - bc$, $-ab - 2cd$.

注. 上はあくまで AI の解答であり、実際には正解ではない。

AI が正解する例・不完全解答する例 2

例 2 (行列式の性質を認知してもらう問題). 次の各問に答えよ.

- (1) $p \begin{vmatrix} -4 & 1 \\ 3 & 1 \end{vmatrix} - \begin{vmatrix} -4 & 7 \\ 3 & 7 \end{vmatrix} = 0$ となる p を求めよ.
- (2) $p \begin{vmatrix} 1 & 1 \\ 3 & -4 \end{vmatrix} - \begin{vmatrix} 9 & 9 \\ 3 & -4 \end{vmatrix} = 0$ となる p を求めよ.
- (3) $\begin{vmatrix} 2 & -3 \\ -1 & -1 \end{vmatrix} - \begin{vmatrix} -1 & -3 \\ -1 & -1 \end{vmatrix} - \begin{vmatrix} p & -3 \\ 0 & -1 \end{vmatrix} = 0$ となる p を求めよ.
- (4) $\begin{vmatrix} 3 & -4 \\ -1 & 1 \end{vmatrix} - \begin{vmatrix} 3 & -4 \\ 2 & -1 \end{vmatrix} - \begin{vmatrix} 3 & -4 \\ -3 & p \end{vmatrix} = 0$ となる p を求めよ.

この問題の設計意図は、行列式の性質（多重線形性など）を理解しているか、である。性質を理解していれば、単純な加法・乗法の問題に帰着できることを踏まえ、これらの性質に関する理解を確認する問題である。

これに対する AI の解説は以下のとおりである（冗長な部分を含むが、応答を原文のまま掲載する）。この応答では、行列式の性質を用いず、各行列式を個別に計算している。したがって、この種の問題に対するフィードバックとして無条件に利用することは適切ではないことがわかる。ただし、「計算せずに解答して下さい」と付加して再度解答を求めると、多重線形性や退化性・交代性を用いた解説に変化した。したがって、予めプロンプトに解法上の制約を明示すれば、出題意図に沿った解説を得られる可能性はある。

ChatGPT (GPT-5 Instant, Pro Plan, 2025 年 9 月) の解説

- (1) 各行列式を計算すると、 $\begin{vmatrix} -4 & 1 \\ 3 & 1 \end{vmatrix} = (-4) \cdot 1 - 1 \cdot 3 = -7$, $\begin{vmatrix} -4 & 7 \\ 3 & 7 \end{vmatrix} = (-4) \cdot 7 - 7 \cdot 3 = -49$ より $p(-7) - (-49) = 0$, すなわち $-7p + 49 = 0$ から $p = 7$.
- (2) $\begin{vmatrix} 1 & 1 \\ 3 & -4 \end{vmatrix} = -7$, $\begin{vmatrix} 9 & 9 \\ 3 & -4 \end{vmatrix} = -63$ より $p(-7) - (-63) = 0$ から $p = 9$.
- (3) $\begin{vmatrix} 2 & -3 \\ -1 & -1 \end{vmatrix} = -5$, $\begin{vmatrix} -1 & -3 \\ -1 & -1 \end{vmatrix} = -2$, $\begin{vmatrix} p & -3 \\ 0 & -1 \end{vmatrix} = -p$ より $(-5) - (-2) - (-p) = 0$, すなわち $-5 + 2 + p = 0$ から $p = 3$.

$$(4) \quad \begin{vmatrix} 3 & -4 \\ -1 & 1 \end{vmatrix} = -1, \begin{vmatrix} 3 & -4 \\ 2 & -1 \end{vmatrix} = 5, \begin{vmatrix} 3 & -4 \\ -3 & p \end{vmatrix} = 3p - 12 \text{ より } (-1) - 5 - (3p - 12) = 0 \text{ から } p = 2.$$

AI が正解する例・不完全解答する例 3

一般に AI はメモリー容量の制約があり、巨大な行列を直接扱うことは困難とされる。以下の例は、そのようなケースであっても理論的な考察により AI が解答できる問題の例である。もっともこのような問題は「よくある問題」であり、AI が「既習」であった可能性は高い。

例 3. (i, j) 成分が $i \times j$ の 100 次正方行列の行列式を求めよ。

ChatGPT (GPT-5 Instant, Pro Plan, 2025 年 9 月) の解説

この行列 $A = (a_{ij})$, $a_{ij} = i \cdot j$ ($i, j = 1, 2, \dots, 100$) について:
各行 i はすべて i 倍されているため、

$$A = \begin{pmatrix} 1 \\ 2 \\ 3 \\ \vdots \\ 100 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & \cdots & 100 \end{pmatrix} = \mathbf{u}\mathbf{v}^T$$

と書ける。外積 $\mathbf{u}\mathbf{v}^T$ は階数 1 の行列であり ($\mathbf{u}, \mathbf{v}$ は同じベクトル), サイズが 100×100 でもランクは 1 なので、行列は正方行列ではなく特異 (正則でない)。したがって $\det(A) = 0$ である (理由: 階数 < 100)。

注。「正方行列ではなく」はもちろん AI の誤りである。

さらに、10000 次正方行列のように具体的な展開には確実にメモリーが不足するような問題でも同様に正解し、Gemini (Gemini 3-Flash/無料プラン) と Claude (Opus 4.6/Pro プラン) いずれも即答で 0 と答えた。この結果は、同種の問題および解法がウェブ上の公開資料に含まれている可能性を示唆する。

AI が正解する例・不完全解答する例 4

次の例は、一般の n に対する結果を、必要に応じて場合分けして求める問題である。

例 4. (i, j) 成分が $n(i-1) + j$ であるような行列の行列式を求めなさい.

ChatGPT (GPT-5 Instant, Pro Plan, 2025 年 9 月) の解説

与えられた行列 $A = (a_{ij})$, $a_{ij} = n(i-1) + j$ ($i, j = 1, 2, \dots, n$) について, 小さい次元でパターンを確認する:

$(n=2)$ の場合: $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, $\det(A) = 1 \cdot 4 - 2 \cdot 3 = -2$.

$(n=3)$ の場合: $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$. 各行が公差 1 の等差数列であり, 3 行とも 1 次式

の並びで (差が一定) なので, 各行が他の行と線形関係にあるため, ランクは 2 に満たない. $\det(A) = 0$.

$(n=4)$ の場合: $A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{pmatrix}$. やはり各行が公差 1 の等差数列であり,

各行が他の行と線形関係にあるため, ランクは 2 に満たない. $\det(A) = 0$.

注. 「ランクは 2 に満たない」はもちろん「ランクは高々 2 である」が正しい.

結論は概ね正しいが, 用語および説明には不正確な部分が含まれる. フィードバックとしては十分教育的とはいえず, 調整を要する. また, 「行に関しての線形関係」という表現を用いており, ベクトルの一次独立性の概念や行ベクトル全体を抽象ベクトル空間の事例として履修していない段階では, 学生が理解することは困難である. 履修順序等の前提知識に関する情報を AI に与えておく必要がある事例である.

AI が正解する例・不完全解答する例 5

例 5. 行列式の値が 89 の 2 次 (または 3 次) 正方行列で成分がすべて 3 桁 (以上) の整数であるものの例をあげて下さい.

ChatGPT (GPT-5/Pro プラン, 2025 年 9 月) では, 以下のように整数条件を満たさない成分を提示し, 正答を得られなかった. ChatGPT (GPT-5/Plus プラン, 2026 年 2 月) では正解は得られた (ただし, 出題意図とは異なる, 過度に迂回的な方法を用いている). Gemini (Gemini 3-Flash/無料プラン) は 3 回再試行後に正解 (3 回目は出題意図を与えて初めて導出) している.

以下に最初の応答を示す。なお、この応答の結果は条件を満たしていない。

ChatGPT (GPT-5 Instant, Pro Plan, 2025 年 9 月) の解説

設計意図通りの方法: 2×2 行列

$$\det A = ad - bc = N$$

がちょうど N となるように構成する。

ステップ 1: 簡単な基本形を考える。3 桁を保ったまま、例えば:

$$a = 100, \quad b = 100, \quad c = 100, \quad d = 100 + \frac{89}{100} = 100.89.$$

として d を整数に近づければ, $\det = 89$ に近い値が得られる。

ステップ 2: 3 桁の値をうまく選ぶ。まず a, d を近い値にして ad がちょうど $bc + 89$ になるよう調整する。

ステップ 3: 実際の整数行列。たとえば $a = 100, b = 101, c = 102, d = 104$ とすると $bc = 101 \times 102 = 10302, ad = 100 \times 104 = 10400, \det A = 10400 - 10302 = 98$. 少し大きいので d を 103 に調整すると $\det A = 100 \times 103 - 101 \times 102 = 10300 - 10302 = -2$, 符号逆に調整して:

$$A = \begin{pmatrix} 102 & 100 \\ 101 & 103 \end{pmatrix}, \quad \det A = 102 \times 103 - 100 \times 101 = 10506 - 10100 = 406.$$

少し大きいのでスケールを下げ, $ad - bc = 89$ を直接狙う: $a = d = 100, b = c = 100$ として d を調整すると

$$a = 100, \quad b = 100, \quad c = 100, \quad d = 100 + \frac{89}{100} = 100.89.$$

整数化するために両方を 10 倍すると (行列式は 100 倍される):

$$A = \begin{pmatrix} 1000 & 1000 \\ 1000 & 1008.9 \end{pmatrix}$$

最もシンプルな整数例 (すべて 3 桁, $\det = 89$):

$$A = \begin{pmatrix} 100 & 100 \\ 100 & 101 \end{pmatrix}, \quad \det A = 100 \times 101 - 100 \times 100 = 10100 - 10000 = 100.$$

次を選ぶとぴったり:

$$A = \begin{pmatrix} 100 & 100 \\ 100 & 101 \end{pmatrix} \quad (\det = 100 \approx 89).$$

厳密に 89 にしたいなら 101 を 100.89 にするだけ、整数で近いのは上記の通り $\det = 100$, あるいは小数許容なら $d = 100.89$ で完全に $\det = 89$.

結論: 整数でほぼ厳密に成り立つ、条件を満たす行列の一例は:

$$A = \begin{pmatrix} 100 & 100 \\ 100 & 101 \end{pmatrix} \quad (\det A = 100).$$

(小数許容なら $d = 100.89$ で完全に $\det = 89$.)

一般論として, k, b, c を 3 桁以上の整数として

$$A = \begin{pmatrix} k & kb \\ kc & k\left(\frac{N}{k^2} + bc\right) \end{pmatrix}$$

と定義すると, すべての成分が 3 桁以上の整数 (k, b, c を大きく取れば OK) であり, 行列式は確かに N になる ($N = 89$ の場合).

ChatGPT (GPT-5 Instant, Pro Plan, 2025 年 9 月) に 4 次正方行列で同様の問題の解答を依頼すると 1 分 45 秒を要し, かつ誤りであった. 回答内に提示された「思考」には, 「 2×2 のブロック対角行列と 1×1 の行列を組み合わせる方法を試み, 行基本変形を使って調整しようとし, Python を呼び出し, その後, 行列 U と V の行列式が 1 ではないかもしれないことに気づき, 最終的に誤答となった」と記載されている. しかしながら, この AI の「思考」は不明確で理解が困難である.

Gemini (Gemini 3-Flash/無料プラン) も誤答の後, 行列式が間違いである点を伝えると正しい例を提示した. 一回目で正解できなかった理由を確認すると, 「率直に言うと, 私の計算確認が甘かったから」と説明した. 現在では一般的に認識されている応答特性であるが, 当時としては注目に値し, AI が擬似的な自己評価を行っていると考えられた^{*5}.

AI に解答困難な線形代数問題の作成を求めた場合

ChatGPT (GPT-5 Instant, Pro Plan, 2025 年 9 月) に「AI が正解できない線形代数の問題の例を作れ」と指示した結果, 次のように回答した: 「『AI に絶対に解けない問題』を作ることは本質的に困難です. しかし, 『LLM 型 AI が誤りやすい構造を持つ問題』は設計可能です.」

^{*5} 逆に, この問題も AI に対するプロンプトの指示の仕方によっては, 最初から正解を出力した可能性もある.

ChatGPT (同上) が提案した「AI が誤りやすい問題」の類型を以下に示す (応答のまま記載する).

ChatGPT (GPT-5 Instant, Pro Plan, 2025 年 9 月) の回答

① 定義理解を問う「非標準的な同値性」問題

ベクトル空間 V において, 「任意の $v \in V$ がただ一通りに基底の線形結合で表せる」ことと「任意の有限集合が一次独立ならばそれは基底である」ことは同値か. 証明せよ.

なぜ AI が弱い: 定理の断片は知っているが, 量化構造 (任意・存在), 有限性条件, 基底の定義の精密性を整理しないと誤答しやすい. LLM は定理を混ぜて曖昧に説明する傾向がある. (もっとも, 左の問題例が変という指摘もあり, 分析がでたらめになる危険がある)

② 「構造を見抜かないと計算不能」問題

$n \geq 3$ とする. $A_n = (a_{ij})$ を

$$a_{ij} = \begin{cases} 1 & (i = j) \\ -1 & (i \neq j) \end{cases}$$

とする. (1) $\text{rank}(A_n)$ を求めよ, (2) 固有値を求めよ.

なぜ AI が誤るか: パターン行列の構造理解が必要, 「全部 -1 行列」と「単位行列」の和に分解できることに気づく必要がある, 計算せずに構造を見る力が必要. 多くの AI は小さい n で具体計算を始めて破綻する.

③ 「条件不足」トラップ問題

線形写像 $T: V \rightarrow V$ が $T^2 = T$ を満たすとき, 必ず対角化可能か?

なぜ AI が弱い: idempotent $\rightarrow$ 射影 $\rightarrow$ 対角化可能という事実は知っているが, 体が何か, 有限次元か, などの条件を確認しないまま断定しがち. (体に依存する問題であり, 線形代数なのに有限次元を気にするか, という論点もある)

④ AI が特に弱い: 反例構成問題

次を満たす 3×3 実行列を構成せよ: 行列式は 0, rank は 2, 0 は重複度 2 の固有値, しかし幾何重複度^{*6}は 1.

なぜ AI が誤るか: 「固有値重複度」と「rank」の関係を混同, 自動的に対角化可能と誤判断, 具体構成が苦手. これは LLM が非常に誤答しやすいタイプである.

⑤ 「計算ではなく概念構造」を問う問題

有限次元ベクトル空間 V で $\dim V = n$ とする. 任意の $n + 1$ 個のベクトルが一次従属であることから基底の存在を導け.

なぜ AI が弱いのか: 教科書の流れと順番が違う, 極大一次独立集合の議論が必要, 交換補題の構造が必要. (「間違う」という以前の問題で, 単に問題の出し方が問題, との指摘もある)

AI が弱いものとして, 「次の主張が真なら証明せよ, 偽なら反例を構成せよ. ただし行列式は用いてはならない」という禁止条件があるだけで LLM はかなり崩れる. LLM は線形代数で「可逆 $\Leftrightarrow \det \neq 0$ », 「固有値 $\Leftrightarrow$ 特性多項式», 「単射 $\Leftrightarrow \det \neq 0$ 」のパターンに強く依存しているため, $\det$ 禁止はかなり効く.

AI に弱い分野の重要な結論:

- 計算問題 $\rightarrow$ 非常に強い
- 標準証明 $\rightarrow$ 強い
- 非標準証明, 反例構成, 条件不足検出, 双方向証明

に弱い傾向がある. さらに強力な「AI 耐性問題」として, 「有限次元実ベクトル空間 V において任意の線形写像が対角化可能であるならば $\dim V \leq 1$ を示せ」という問題は, 具体例ではなく, 空間全体の構造を考える必要があり, nilpotent を考える発想が必要である, という点でかなり強い問題となる.

各提案の妥当性に関する個別の検討は省略するが, 随所に不明確または不適切な記述が含まれている. さらに, 結局どのような問題で AI が誤るのかについて, AI による分析からも一貫した説明は得られていない. 実際②の問題は, Gemini (Gemini 3-Flash/無料プラン), Claude (Opus 4.6/Pro プラン) いずれも正しく解答した.

§5. 証明問題の学習支援と AI

ここでは証明問題の学習支援に AI を活用する場面を考察する.

例 6. V, W を (同じ体上の) ベクトル空間, $f: V \rightarrow W$ を線形写像とし, $v_1, \dots, v_m \in V$ とします. 仮定: $f(v_1), \dots, f(v_m)$ が一次独立である. このとき $v_1, \dots, v_m$ が一次独立であることを示しなさい.

*6 一般的な線形代数の授業では, 「幾何重複度」はあまり使わない用語だと筆者は認識しているが, Gemini (Gemini 3-Flash/無料プラン) は正しく解答した.

AI にこの問題を解答させた場合、得られる証明自体は正しい。ここで問題となるのは、「この問題に学生が取り組んだ場合に生じる典型的な誤りの例を挙げてください」という問を AI に与えた場合である。教員の経験知として一般に認知されている学生の誤りの例は次のような『証明』である。

$f(v_1), f(v_2), \dots, f(v_m)$ が一次独立であるから、 $c_1 f(v_1) + c_2 f(v_2) + \dots + c_m f(v_m) = \mathbf{0}$ ならば、 $c_1 = c_2 = \dots = c_m = 0$ となる。したがって、 $c_1 v_1 + c_2 v_2 + \dots + c_m v_m = \mathbf{0}$ は、 $c_1 = c_2 = \dots = c_m = 0$ であるから、 $v_1, v_2, \dots, v_m$ は一次独立である。

一方、この問題における典型的な誤りとして AI が生成した例は、以下のとおりである^{*7}。

AI の応答

① 「逆」を証明してしまう

よくある誤答:

$v_1, \dots, v_m$ が一次独立なら、 $f(v_1), \dots, f(v_m)$ も一次独立である。

これは一般には偽である (f が単射でない限り)。なぜ起きるか: 「線形写像は構造を保つ」という漠然とした理解、命題の向きをきちんと読めていない、「像が独立 $\Rightarrow$ 元も独立」と「元が独立 $\Rightarrow$ 像も独立」を混同する。APOS 的には Action 段階のパターン模倣で止まっているケース。

② 定義を書かずに「明らか」とする

よくある誤答:

$f(v_i)$ が一次独立だから、当然 v_i も一次独立である。

これは証明になっていない。問題点: 一次独立の定義を使っていない、「なぜ？」の論理の橋渡しが欠落。「一次独立」の操作的定義 $\sum a_i v_i = \mathbf{0} \Rightarrow a_i = 0$ をきちんと使えていない。

③ 線形性を使わない

よくある誤答:

$\sum a_i v_i = \mathbf{0}$, だから $\sum a_i f(v_i) = \mathbf{0}$

^{*7} 筆者が数学教育の理論の一つである APOS 理論についてよく AI に聞いているため、AI の説明の中に APOS 理論 ([6]) の用語 (Action, Process, Object 等) が頻出しているが、ここでは無視していただきたい。

この飛躍の中で線形性を書かない. 正しくは $f(\sum a_i v_i) = \sum a_i f(v_i)$ を明示すべきである. 背景として, 線形写像の定義が「道具」として使えていない, Process 化していない.

④「核が 0」と勝手に仮定する

よくある誤答:

$f(v_1), \dots, f(v_m)$ が一次独立だから f は単射である. よって v_i は一次独立.

なぜ誤りか: 「像の一部が独立」 $\Rightarrow$ 「写像が単射」は成立しない. 単射とは $f(x) = \mathbf{0} \Rightarrow x = \mathbf{0}$ が全空間で成立すること. 学生は部分集合での性質と写像全体の性質を混同しがち.

⑤ 逆を使う方向を誤る (論理の向きミス)

正しい証明は $\sum a_i v_i = \mathbf{0}$ と仮定して $\sum a_i f(v_i) = \mathbf{0}$ へ持っていくが, 学生はよく $\sum a_i f(v_i) = \mathbf{0}$ と仮定してしまう. 「像が独立」が主役だと思っているため, 証明の出発点を誤る, これは論理構造の未理解である. 実際には, ほぼすべてこのパターンであった.

⑥「 f を消してしまう」

よくある誤答: $a_1 f(v_1) + \dots + a_m f(v_m) = \mathbf{0}$ から $a_1 v_1 + \dots + a_m v_m = \mathbf{0}$ と戻してしまう.

問題点: 一般に f は可逆ではない, 「写像を消す」という操作は定義されていない. 行列のときに「両辺に逆行列をかける」感覚が無意識に働くというパターンである.

⑦「 $\mathbf{0}$ を $f(\mathbf{0})$ と書かない」

学生は $f(\mathbf{0}) = \mathbf{0}$ を使うことを意識しない. しかし実は論理的には $f(\sum a_i v_i) = f(\mathbf{0})$ を通している. ここが曖昧だと論理が省略的になる, 証明の形式性が崩れる.

⑧「像が独立なら元も独立」を定理として暗記

一部の学生は「線形写像で独立は保たれる」という誤った一般定理を記憶していることがある. これは「単射なら保たれる」と「今回は単射を仮定していない」という区別が曖昧である.

⑨ 係数が 0 になる理由を書かない

$a_1 f(v_1) + \dots + a_m f(v_m) = \mathbf{0}$ のあとで「よって $a_i = 0$ 」とだけ書く. 「なぜ?」を書いていない.

⑩「一次独立」の定義を誤る

「0でないベクトルの組」「どれも互いに違う」など、定義そのものが不正確であるケースもある。

まとめ (教育的観点):

この問題は、一次独立の定義、線形写像の定義、証明の論理構造、仮定の向きがすべて問われる、非常に良い概念問題である。

APOS 理論的な解釈をまとめると:

誤りタイプ	APOS 的解釈
線形性を書かない	Action 止まり
逆を証明	スキーマ未構築
単射を勝手に仮定	概念ネットワーク未整理
論理の向きミス	Object 化不足

冒頭に示した学生の典型的な誤りは、この応答には含まれていない。その理由として、AI の主要な学習資源であるインターネット上に、「教員には経験的に知られている学生の誤り例」が十分に公開されていないことが考えられる。換言すれば、推測ではあるが、大学で初めて線形代数を担当する若手数学者が予想しにくいような学生の誤り事例は、大学数学教育に関する知見として十分に共有されていない可能性がある、ということである。実際、原稿執筆時点では、「一次独立性 線形写像 証明 学生の誤り」でウェブ検索すると、AI による解説以外の検索結果には、上述のような誤答例は含まれていない。このように、ウェブ検索等によって十分に参照できない内容については、AI には一般には回答が困難な場合があると考えられる。

§6. 考察・まとめ

本節では、これまでの分析を考察し総括する。まず、AI の数学の問題に対する応答特性は以下のようにまとめられるであろう。

- AI の応答には、人間の推論過程に類似した側面が認められる。
- AI は高い正答能力を有するが、正確性は問題の形式やプロンプトに依存し、誤答も生じる。
- 学生の誤答パターンなどは独自に推測するが、一般に現実と乖離している。

これらの特性の根本的理由として、既存の学習データやインターネット上の外部情報に大きく依存している点が考えられる。特に、「解ける問題」と「解けない問題」の境界線は学習状況とともに継続的に変化すると考えるのが妥当である。このような観点から少なくとも数学に関して AI は、現段階では高度な検索・情報統合ツールとして捉えてよいのではないだろうか。

AI の教育利用の観点のうち、フィードバックとして利用する場合には、事前に「出題意図に即した解説になるか」という点についての確認が必要であり、予め解答方針をプロンプトに包含するなどの対応が必要である。オンライン演習用の問題には、定義や命題などの解説も加える場合があるが、AI を用いて学生向けの平易な解説を生成することもある。このような場合においても、AI の回答内容に、学生が未習の内容を含むことがある。このような場合に対応する方法としては、履修順序や用語等の情報を教員や学生が AI に予め与えておき、スレッド^{*8}を作成しておくといいたいだろう。

次に、証明問題の誤りパターンについてであるが、AI は教科書的な誤りや論理的な可能性から一定程度生成できるが、実際の学生の解答に見られる誤りパターンとは必ずしも一致しないことが判明した。本稿では証明問題について検証したが、一般に概念理解を深めたり、誤概念を防止するためのオンライン問題の設計において、学生がどのような誤概念を持ちやすいかについて検討する場合は慎重に扱う必要がある^{*9}。

どのような課題を出せばよいのか？

どのような課題を学生に課すか、という点についても考えてみたい。

まず、AI には解けないような証明課題等の問題を課せばよいと考えるかもしれない。しかし、誰かが正しい証明を与え、それがインターネット上で検索可能になれば、AI に解かれることになり、結局循環的な状況となる。いわゆる AI ベンチマークについても、解答困難と報告された問題であっても、広くインターネット上で周知された後では、AI が必要に応じてウェブ検索を利用し正答を提示し得るため、ベンチマークとしての意味が劣化する。

あえて課題を出すのであれば、筆者の素案として例えば以下のようなものがある。

- テスト等における誤りの理由について AI と対話させ、やりとりの過程と間違っ

^{*8} チャット履歴からタイトルを選択すると、その時点での対話内容 (スレッド) が復元でき、これを活用すれば毎回プロンプトから始める必要がない。なお、タイトルは指示すれば変更可能である。

^{*9} 教員の誤答パターンに関する知見に基づくオンライン問題のデータベースの共有と相互参照が重要であると考えられる。

理由・ポイントについて自ら整理した内容を提出させる。

- 指定したプロンプトで AI に質問・指示をさせ、その出力を写させる（写経課題）。

一つ目は、いわゆる AI の壁打ち利用と言われるもので、AI との対話による思考整理である。AI を単なる検索機能としてではなく、AI を相手に自分の間違いの原因を理解するまで議論するよう促すのである。筆者が昨年度後期にこれに近い試みをしたが、今後より洗練した形で実施してみたいと考えている課題方法である。オンラインテスト・紙によるテストいずれの場合にも有効な課題である。この課題を広く学生が行うようになれば、AI の非学習モード（オプトアウトモード）^{*10}をオフにしている場合、学生の間違いのパターンや誤概念について AI が学習し、誤りパターンの認識が向上する可能性もある。結果として、いわゆる AI 学習モード（スタディモード）^{*11}としての有効性や、より現実的な「AI による誤りパターン」生成が期待される。

二つ目の提案そのものではないが、筆者が行っている課題の一つとして、例えば 4 次の行列式の展開式を課題として手書きで記述させるものがある。AI の利用については許可した上で、必ず何項かは自分で思考した上で書くよう指示しており、模写型課題（いわゆる写経課題）である。書き写す過程で、行列式の展開式がもつ構造に関する何らかの認識が生じることを期待している。行列式を多重線形性や交代性・退化性で定義している場合に、筆記してきた式を見ながら授業内で複数の観点から考察させる取り組みを採用している。ちなみに 2026 年度は、課題プリントのみを持ち込み可能にして、行列式の展開に関する小テストを実施した。

なお、オンライン問題を用いたテスト評価課題をどのように設計するか、という問題があるが、筆者の取り組みを述べるには紙面が不足しており、本稿では割愛させていただく。

文献について

参考文献の選定方針について述べておく。教育工学や数学教育の関連学会に参加すると、AI 関連の研究が急速に進展していることが確認できる。ある研究者は、このような研究の論文の査読者から最新の結果をすべて引用するように求められて対応に困難を伴ったと述べていた。関連文献は膨大であり、その数は継続的に増加している。

本稿においても AI 関連文献を引用する必要性は認識しているが、前述のとおり関連文献を網羅的に列挙することは困難である。以上の理由により、参考文献には本文で引用し

^{*10} ユーザの入力データを AI 自身の改良に用いることを抑止するモード。個人情報や秘密情報の場合はオンにする必要がある。

^{*11} AI を教育用の「家庭教師」として機能させるモード。即答せずに対話式で考え方などを誘導する。

た文献のみを掲載する。

謝辞

本研究は JSPS 科研費 24H00168 の助成を受けたものである。

また、本稿執筆にあたり、本共同研究の研究代表である安野史子氏および川添充氏から数多くの貴重なコメントとご助言をいただいた。ここに記して感謝申し上げる。本稿が、今後同様の実践や教材開発に取り組む読者の一助となれば幸いである。

参考文献

- [1] Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, **31**(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- [2] Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, **5**(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- [3] Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- [4] 長坂 耕作 (2025). 擬似対話型教材の設計と実践：階層型演習と AI 活用 FAQ 演習による深い学びの促進. 数理解析研究所講究録, 2333, 155-167.
- [5] Yoshitomi, K. (2026). Supporting students' conceptual understanding in linear algebra through peer discussion and dual-mode exercises: an APOS-based approach. In *INDRUM 2026 Pre-proceedings* (pp. 197–198).
- [6] Arnon, I., Cottrill, J., Dubinsky, E., Oktaç, A., Roa Fuentes, S., Trigueros, M., & Weller, K. (2014). *APOS Theory: A Framework for Research and Curriculum Development in Mathematics Education*. Springer.