2025-07-13　麻生塾_AI診情の進捗

▼ 目次

これは何？
過去資料
5月振り返り時の議論整理
1. 授業の基本方針
2. 授業の進め方と学生の様子
3. 評価方法と今後の課題
4. 今後の改善方向性
5. 具体の改善詳細と、改善後の評価項目
【改訂版 v2】AI×診療情報授業 評価フレームワーク
Appendix

これは何？

- AI診情の授業を進める中で、いくつかの課題が出ている
- その結果、評価手法の見直しが必要
- 論点と、評価方法の見直しの具体的な方法を整理した資料

過去資料

2025-02-16 【リサーチ】AI人材育成に向けた学科転換_開発モデルの検証について

2025-02-18 【評価骨子案】医療現場学科転換_Ver1

5月振り返り時の議論整理

1. 授業の基本方針

- **講師は教えない**: 講師は知識を直接教えず、学生が自ら学ぶことを徹底
- **ChatGPTが教師役**: 学生はChatGPTとの対話を通じて、プログラミング（Python）の概念（for文、APIなど）を自律的に学習
- **講師の役割**: 学生の学習時間を最大化するため、授業中は基本的に沈黙。ChatGPTにログインできないなど、学習の進行が止まるトラブルのみ積極的にサポート

2. 授業の進め方と学生の様子

- **授業の流れ**:
  1. **初回**: Google Colabの使い方と「変数」など、最低限の基礎知識のみ講師が解説。

2. **通常授業**: 学生は90分間、黙々とPCに向かい、ChatGPTを使って課題に取り組む。
3. **授業の最後（5分）**: 講師が「AI活用事例」や「現役エンジニアの働き方」などのコラムを話し、学生の視野を広げる。

- **学生の様子**:
  - ツールの操作に慣れ、多くの学生が授業時間いっぱい集中して取り組んでいる
  - タイピングスキルも問題なく、講師よりもChatGPTに質問する方が速くて気楽だと感じているよう
  - 当初の想定を大幅に上回る**3〜4倍の速さ**で学習が進んでいる。これは、学生が「ChatGPTにうまく質問するスキル」を早期に習得したためでは？今年のカリキュラムが今の段階で終了しそう

3. 評価方法と今後の課題

- **現状の評価方法**:
  - 学生が作成したコードを、専用の「テストAI（GPTs）」に提出。
  - AIがコードを基に理解度テストを出題し、合否を判定
- **課題**:
  - **真の理解度の把握が難しい**: 実行できないコードでもテストに合格できてしまうなど、評価方法がハックされる可能性が判明
  - **進捗が想定より速すぎる**: 後期用に準備していた課題も前倒しで実施しており、今後の授業計画の調整が必要
  - 授業提供側としては、理解度確認のテストを作ってしまう。カリキュラムの達成目標が明確であればあるほど、AIに聞くと答えが得られてしまう構造が発生
    - 問ができると答えがわかる世界において、一般的なテストでの能力評価が陳腐化していってしまう。
    - AIによりテストがハックされやすい構造が生まれる。
    - プロセスの評価を入れていき、本質的な理解度をチェックしていく仕組みが必要

4. 今後の改善方向性

担当教員で常に振り返りを行い、アジャイル（迅速かつ柔軟）に改善を進めてくれている。

- **評価方法の改善**:
  1. **テストAIの修正**: コードが正しいか最初にチェックし、不正な場合はテストに進めないように改良
  2. **多角的な評価**:
     - 提出物はコードだけでなく、**AIとの対話ログも共有が必要**
     - 授業中に**口頭での質疑応答**を行い、理解度を直接確認
     - 定期的に同じ形式のテストを実施し、**理解度の定点観測**

5. 具体の改善詳細と、改善後の評価項目

1. **成果物主義からの脱却**: 完成したコードだけでなく、**ChatGPTとの対話ログや試行錯誤のプロセスそのもの**を評価の中心に据えていく
2. **「言語化能力」の重視**: AIから得た知識を**「自分の言葉で説明できるか」**を測るため、口頭試問やコードへのコメント記述を必須の評価要素に加えて、本質的な理解度を測る
3. **自律性の評価**: 指示された課題だけでなく、**自ら課題を発見し、探求する姿勢**を評価項目に加えることで、学生のポテンシャルと学習意欲を正しく捉える

【改訂版 v2】AI×診療情報授業 評価フレームワーク

評価項目	S (目標を大幅に上回る)	A (目標水準に達している)	B (目標にあと一歩)	C (目標到達に大きく不足)	D (著しく未達)
1. AI活用による課題解決能力 **(評価資料: 成果物, ChatGPT対話ログ)**	未知の課題に対し、自ら解決への道筋を設計し、AIを「思考のパートナー」として活用できる。対話ログには、高度な指示や多角的な質問が見られ、生成されたコードをさらに改善・最適化している。	提示された課題に対し、AIと適切に対話しながら必要なコードや情報を引き出し、解決できる。対話ログは論理的で、試行錯誤の過程が明確に記録されている。	課題解決は可能だが、AIへの指示が断片的であったり、非効率な対話が多かったりする。生成されたコードをそのまま利用する傾向があり、応用力に課題が見られる。	AIとの対話がうまく成立せず、課題解決に多くの時間を要する、または未達に終わる。エラー発生時に自力で次の質問を組み立てることが困難。	AIツールを操作して課題解決の糸口を見つけることができず、対話ログも提出できない、または内容が著しく乏しい。
2. 技術・概念の言語化・応用能力 **(評価資料: 口頭試問, コードコメント)**	自身のコードの動作原理や採用した技術（API等）について、**背景や代替案も含めて**自身の言葉でよどみなく説明できる。他の学生に教えることも可能。	口頭試問において、自身のコードのロジックや主要な概念を**自身の言葉で正しく説明**できる。コード内にも処理内容を説明する適切なコメントが記述されている。	コードの動作を説明できるが、専門用語の誤用や概念のあやふやな理解が見られる。口頭試問で質問に詰まる場面がある。	コードがなぜ動くのかを説明できず、AIからのコピー＆ペーストに留まっている。口頭試問で応答できない、または根本的な誤解をしている。	自身のコードについて全く説明できず、基本的な用語の理解もできていない。
3. 医療データ活用・分析スキル **(評価資料: 成果物, 対話ログ, 口頭試問)**	データの背景（欠損値の意味等）を考察し、AIと対話しながら**複数の分析手法を比較検討**した上で、適切な手法を選択できる。分析結果を	提供された医療データに対し、AIを活用して適切な前処理と分析を実行できる。口頭試問で、**なぜその処理が必要か、分析結果が何を意**	データ分析の手順は実行できるが、前処理や分析手法の選択理由を十分に説明できない。結果の解釈が統計的な数値を述べるに留まり、医療	AIの支援があっても、データの前処理や分析で多くのエラーを出す。分析結果の解釈に根本的な誤りがあり、対話ログからも試行錯誤の意図	医療データの基本的な扱い方が分からず、分析に着手できない。データの読み込みや中身の確認すら困難な状態。

	医療的な文脈で深く考察し、次のアクションを提案できる。	**味するか**を自身の言葉で説明できる。	的な意味付けが乏しい。	が読み取れない。	
4. 倫理・安全性の実践 **(評価資料: 演習中の行動, 課題の前提条件遵守)**	AI利用の倫理的・法的リスクを深く理解し、演習中に**他の学生が気づかないような潜在的リスクを指摘**できる。個人情報等の扱いに極めて慎重である。	データ匿名化や著作権、プロンプトインジェクション等のリスクを理解し、安全ガイドラインを遵守して演習に取り組める。指摘された点にも適切に対応できる。	倫理・安全上の基本的な留意点は理解しているが、時折、個人情報に関する配慮や引用元の明記などで軽微な不備が見られる。	演習中に不適切なプロンプト（個人情報の入力等）を試みるなど、リスク認識の欠如が見られる。講師からの度重なる注意を要する。	倫理・安全への意識が著しく欠如しており、演習において危険な操作を行う可能性がある。
5. 協働・実践力 **(評価資料: グループワーク, 模擬演習)**	チーム内でAI活用をリードし、自身の知見を他者に共有することで、チーム全体の生産性を高めることができる。模擬演習では、状況に応じた的確な判断・提案ができる。	グループワークにおいて、自身の役割を理解し、学んだAI技術を使って貢献できる。チームメンバーとの報連相も適切に行える。	チーム内での活動にやや消極的で、自身の意見発信や他者への働きかけが少ない。AIの活用も指示待ちになる傾向がある。	グループワークで孤立しがちで、チームへの貢献が限定的。AIの知識を実践の場でどう活かすか、イメージが湧いていない。	チームでの共同作業が困難で、コミュニケーションに課題がある。演習の進行を妨げる場合がある。

前回評価項目

評価項目	**S（目標を大幅に上回る）**	**A（目標水準に達している）**	**B（目標にあと一歩）**	**C（目標到達に大きく不足）**	**D（著しく未達）**
AIリテラシー基礎理解	授業で扱ったAIの概念・原理を確実に理解し、応用例も自主的に調査して発表できる。専門用語の誤用がなく、理解度テストで高得点を収めている。	授業範囲のAI概念を概ね正しく理解している。専門用語の理解も十分で、理解度テストで合格基準を満たしている。	基本的な概念は理解しているが、複雑な部分や詳細に誤解・不明点がある。理解度テストでは惜しくも合格基準に届かない。	授業範囲のAI概念の理解が浅く、主要ポイントに大きな誤解がある。一部は把握しているものの追加学習が大いに必要。	授業範囲のAI概念をほとんど理解できておらず、主要ポイントを取り違えている。再受講レベルのフォローが必要。
AIツールの操作・応用力	演習で使用したAIツールを自在に操作でき、与えられた課題に対して適切な手法を選択・応用できる。トラブル発生時も自力で解決策を探れる。	AIツールの基本操作を習得し、指示があれば課題に沿った活用ができる。標準的な操作には問題ないが、応用には指導を要する場合がある。	AIツールの操作にまだ不安があり、一部の機能で戸惑いが見られる。課題遂行には断続的なサポートが必要。	操作手順を十分に理解しきれず、演習課題の遂行に大きな支障が出ている。継続的なサポートが欠かせない。	ツールの基本操作が全くできず、課題を進められない状態。操作方法の再習得が必須。

医療データの活用・分析スキル	提供された医療データを適切に前処理し、学習済みの分析手法を用いて有意義な結果を導き出せる。結果の考察も論理的に行えている。	学習した分析手法を用いたデータ解析ができ、前処理や基本的な結果解釈も正しく行えている。ただし深い考察はまだ課題。	データ解析の手順自体は理解しているものの、実際の処理や結果解釈にミス・不十分な点が散見される。	データ分析手順の理解が不十分で、結果の解釈にも大きな誤りがある。演習での改善が急務。	データ分析の基礎が確立しておらず、結果を導き出す段階に至っていない。継続的な指導が必要。
倫理・安全意識	ケーススタディやディスカッションでAI利用時の倫理・安全上のジレンマに積極的に発言し、妥当な対応策を提案できる。日頃から安全ガイドラインを遵守している。	一般的な倫理・安全上の留意点（データ匿名化や説明責任など）を理解しており、指摘があれば適切に対応できる。自主的な指摘や深い考察はもう一歩。	倫理面で考慮すべき事項の理解が不十分で、演習中に教員から注意を受けることがある。安全ガイドラインの遵守も一部不徹底。	倫理・安全上のリスクを十分に認識できず、演習やディスカッションでも重大な見落としを指摘される。	倫理・安全に関する意識が著しく欠如しており、プライバシー保護や安全管理に対する理解が全く不足している。
実務適応力（病院実習を含む）	模擬的な臨床シナリオ演習で学んだAI技術を的確に活用し、チーム内で有効にコミュニケーションが取れる。問題発生時にも臨機応変に対処できている。	シナリオ演習やグループワークで、指導のもとAI活用を試み、基本的な現場対応ができる。チームでの報連相も概ね良好。	実践演習でAIを活用する際に戸惑いが多く、自信を持って行動できない。チーム連携にも消極的で、さらなる経験が必要。	演習でAI活用に大きな混乱が生じ、チーム連携にも支障をきたす場面がある。現場対応力の育成が急務。	基本対応が困難で、AI活用以前に問題が多く、指導者の継続的フォローが不可欠。

この評価は、単一のテストではなく、以下ポートフォリオ（複数の証拠）を組み合わせて総合的に判断する方法を採用している

1. **課題提出物**:
    - **成果物（コードなど）**: 課題を達成できたか。
    - **ChatGPT対話ログ**: どのように課題を解決したか。（プロセス）
2. **定期的インタビュー（口頭試問）**:
    - 月に数回、数名の学生を対象に5分程度のインタビューを実施。
    - 「このコードは何をしているの？」「なぜこのAPIを選んだの？」といった質問で、本質的な理解度を確認。
3. **観察・記録**:
    - 授業中の自律的な学びや、グループワークでの協働姿勢を講師が観察し、記録。

Appendix

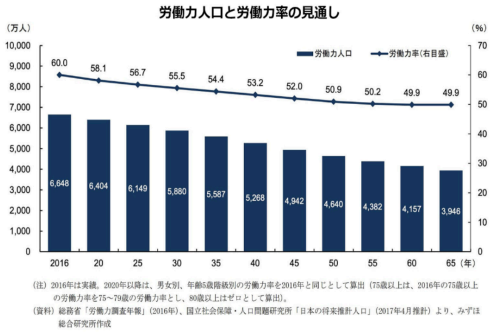
結果ではなく、プロセス及び、意思決定ができる能力や経験が開発されているか？が重要
進行速度を踏まえたカリキュラムの見直しで、エージェント周りのテーマも取り扱えると、生産性改善レベルにあわせた評価指標を追加することが可能なので、引き続き検討を進める

仕事はいつだって奪われてきた

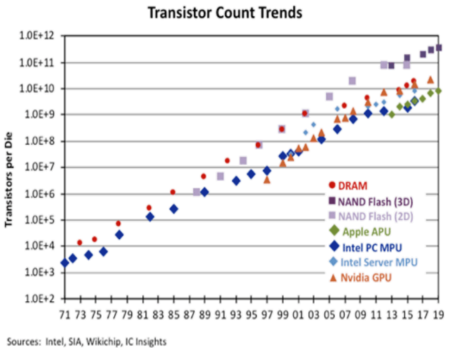
- 技術の進歩は常に仕事を奪ってきたが、それは結局人類が望まない「非人間的」な仕事だった
- 仕事の内容はどの仕事も変わり続けている。
- 新しい技術は必ず新しい仕事を生み出し、人類の生活をより豊かにしてきた。
- **重要なのは新しい技術にオープンマインドで向き合い、学び続けること**

人の価値は高くなり、半導体は安くなり続けている

より少ない労働力で、多くのことをコンピュータに任せる必要



（注）2016年は実績、2020年以降は、男女別、年齢5歳階級別の労働力率を2016年と同じとして算出（75歳以上は、2016年の75歳以上の労働力率を75～79歳の労働力率とし、80歳以上はゼロとして算出）。
（資料）総務省「労働力調査年報」（2016年）、国立社会保障・人口問題研究所「日本の将来推計人口」（2017年4月推計）より、みずほ総合研究所作成



Sources: Intel, SIA, Wikichip, IC Insights

かつてメンバーが提案し、マネージャが決める。

課題を整理し、意思決定し、環境を整える。

メンバーの仕事：実装したり提案する



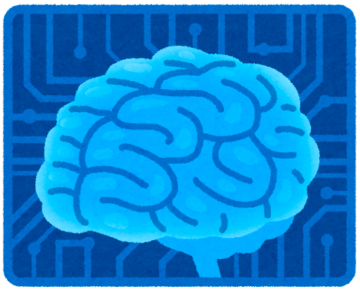
マネージャの仕事：目的の提供と意思決定



今はAIが提案し、人間が決める

課題を整理し、意思決定し、環境を整える。

LLM の仕事：実装したり提案したり



人間の仕事：目的の提供と意思決定



生成AIによる生産性改善のレベル分け

できる化から自働化、自創化をめざしていく。

Level 1	Level 2	Level 3	Level 4	Level 5
AIチャット利用	専門家ボット作成	社内エージェント化	自動化ワークフローへの組み込み	知識創造プロセスエージェント

AI時代の仕事は意思決定で「高密度化」した

- よりよい設計を提示しないといけない。
- 複数のエージェントに仕事を振り分けないといけない。
- 新しい要件を次から次に提案しないといけない。
- メンバーが5人のチームリーダーになるような変化が起きている。

AIエージェント時代には効率化の視点が変わる

昔は、標準化やマニュアル化、最近は可視化と最適化。「正しさ」の視点が変わる。

- **できる化**(頼む)
  - 人がやっていた仕事を「自分でできるようにする」
  - 自分がやっている仕事を人ができるようにする。
- **自働化**(任せる)
  - ワークフローやAgenticな仕組みで仕事をAIに任せる
  - 不確実性の発生源を捉えて自働適応する仕組み
- **自創化**(指し示す)
  - 仕事を回すフェーズではなく仮説検証し、創造するフェーズも自働化する