CN-002

AI 倫理に関するケースメソッドの効果測定とディスカッション分析

Effectiveness measurement and discussion analysis of case methods related to AI ethics

細谷慶人† 家入祐也†‡ 菱山玲子†
Yoshihito Hosoya, Yuya Ieiri, Reiko Hishiyama

1. はじめに

近年，ディープラーニングの開発などにより，AI の技術が飛躍的に進歩している．それに伴い，AI が引き起こす倫理的な問題が注目されており，この現状から，民間人へのAI リテラシー教育が求められている．AI 倫理問題についてのリテラシー教育は，ディスカッション形式などを取り入れ，自ら考えさせることが重要だとされており[1]，海外の MBA(経営学修士)などで用いられているケースメソッドはこれと親和性があると考えられる．

そこで本研究では，AI 倫理に関するケースメソッドを実施し，その教育効果を測定し，ディスカッション内容を分析した．

2. 先行研究

文献[2]では，中学生への AI リテラシー教育を実施し，効果を分析している．この研究は，AI の開発や使い方にフォーカスしており，倫理問題には着目していない．文献[3]ではケースメソッドのディスカッション内容を分析することにより，異なる母国語の被験者間の文化的差異を分析している．この研究で用いられているケースメソッドとは，ケースメソッド教育とは，「討議用ケース」を用いて行う討議型授業をつなげてカリキュラムを構成していく教育形態の総称 [4]と定義されている．文献[3]はケースメソッドのディスカッション内容を分析するという点で，共通点がある．文献[5]では，農業者教育という分野のケースメソッドの適用を試みているが，AI 倫理教育に対するケースメソッドの適用はいまだに少ない．

3. 提案

本研究では，ＡＩ倫理に関するリーディングケースを執筆し，実験参加者間でケースメソッドを行い，そこから収集したテキストデータを分析することにより，ケースメソッドの教育効果を明らかにした．また，ディスカッション中の発言を，文系・理系，男性女性というセグメントで分析することにより，各セグメントのディスカッション中の振る舞いの違いを明らかにする．

本研究に先立ち設計したシステム構成図を図 1 に示す．実験参加者は各インターフェースから，ケース及び事前問題共有 DB とチャット DB にアクセスする．このシステムにより，参加者はチャットベースでリモートのケースメソッドを行うことができる．

以下の 5 つの仮説を立てた．

† 早稲田大学大学院 創造理工学研究科
Department of management engineering, Waseda University
‡ 日本学術振興会特別研究員（PD）
Japan Society for the Promotion of Science, Research Fellow (PD)

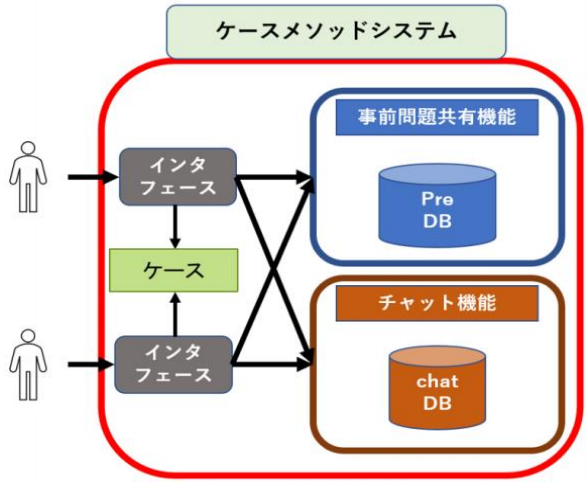

図 1 実験システム概要図

仮説 1：事前問題回答・事後問題回答の間で問題解決方法への言及率が向上する．
仮説 2：データの代表性やバイアスについての議論は理 系の方が活発化する．
仮説 3：透明性やアカウンタビリティなどの責任問題についての議論は文系の方が活発化する ．
仮説 4：ジェンダー問題についての議論は女性の方が活発化する．

仮説 1 はこのケースメソッドにより，問題の理解が深まり，解決案を記述できるようになるという仮説である．仮説 2 及び仮説 3 は，文献[6]より，問題解決において文系学生と理系学生で着眼点が異なることがわかっており，仮説を立てた．仮説 4 は，文献[7]より，中学校教員において女性の方がジェンダーフリーの傾向が強いということがわかっており，仮説を立てた．

4. 実験概要

実験は全て二人一組のペアで行う．被験者は大学生の男性 7 名 (うち文系 3 名，理系 4 名)女性 7 名(うち文系 4 名，理系 3 名)であり，表 1 にその組み合わせを示した．

一般的なケースメソッドの授業は次のように行われる．まず，参加者は予習として与えられたケースに一通り目を通し，それらについての論点を整理し，考えをまとめる．そして，ケースに対する考えを被験者間で共有し，ディスカッションを開始する．以上を踏まえ，2020 年 11 月 28 日から 2020 年 12 月 6 日にかけて，以下のような手順で実験を行った．以下に各手順の詳細を示す．

手順 1．ケース予習・事前問題回答(時間：20 分)
参加者はケース教材を一読後，事前問題についての回答を作成する．事前問題は「文章から考えられる現状の問題点を自分なりに考えてみてください．可

表 1 実験パターン一覧(単位：組)

			参加者				計
			文系		理系		
			男性	女性	男性	女性	
参加者	文系	男性	0				0
		女性	0	1			1
	理系	男性	1	1	1		3
		女性	2	1	0	0	3
計			3	3	1	0	7

能であればその解決策も記述してください」の文章を出題した．参加者は回答をウェブサイト上の入力フォームに自由記述形式で入力する．以下より，この回答を「事前問題回答」とする．

手順 2．ディスカッション(時間：40 分)

簡単なアイスブレイクの後，参加者間でケースについてのディスカッションを行う．ディスカッションは，ファシリテータ等は参加せず，参加者間のみで行う．

手順 3．事後問題回答(時間：10 分)

ディスカッション終了後，事前問題と同じ問題文を出題し，入力フォームに自由記述形式で入力する．以下より，この回答を「事後問題回答」とする．

実験に先立ち，ケースメソッドを行うためのウェブ上のシステムを開発した．このシステムにより，参加者は文字ベースでケースメソッドを行うことができる．図 2 にユーザインタフェースを示す．

実験に先立ち，「エントリーシート自動採点 AI による弊害」というタイトルの 1400 字程度のケースを執筆した．このケースは，2017 年，Amazon 社[8]が採用活動用 AI ツールから女性差別を取り除くことができず，同プロジェクトを中止したという事例[9]があり，このようなシステムが利用され実際に不採用となった女子大生についての内容である．ケースから誘発したいと考えた議論のポイントとしては

1. 学習データの代表性から生じるバイアス．
2. アカウンタビリティなどの企業が果たすべき責任問題．
3. 研究チームの男女比から生じる問題．

の 3 点である．

以下にストーリーの要約を示す．

(ストーリー)

佐藤美咲は理工学部の四年生で，IT 業界を志望していた．佐藤の第一志望の田中総研は AI の研究に力を入れており，その一環として就活生のエントリーシートを AI を用いて分析し，その会社への適正率を算出し合否の決定材料として使用していた．一次選考の合否は不合格．佐藤は落胆した．佐藤は田中総研に AI のアルゴリズムの説明を要求した．

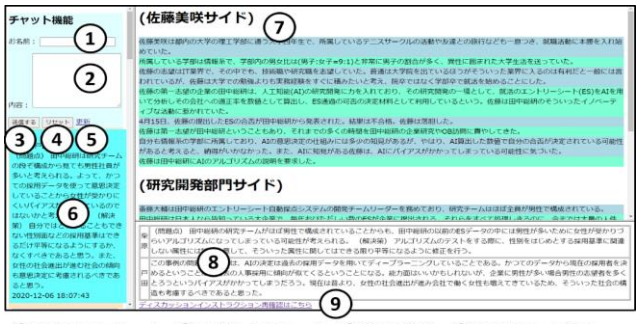

図 2 ユーザインタフェース

5. 実験結果

実験から事前問題回答，事後問題回答及びディスカッションログを取得できた．ディスカッションログの概要を表 2 及び表 3 に示す．ここで，ディスカッションログとは，参加者が一度のチャットに送信した文章を示し，文とはディスカッションログを改行と読点で句切った文字列示し，7 組の平均はそれぞれ 7.86，14.79 であった．本章では実験の分析結果を示す．

5.1. 事前問題回答及び事後問題回答の分析

まず，本実験での教育効果を明らかにするため，事前問題回答及び事後問題回答の分析を行った．分析では初めに，事前問題回答及び事後問題回答を

1. 問題点のみに言及している
2. 問題点及び解決案に言及している

の二つの属性に分け，分類を行った．

なお分類は，著者と協力者(実験の被験者ではない)の二名で行い，お互いに分類終了後 kappa 係数(k)[10]を用いて信頼性を評価した．分類の結果 k=1 で完全に一致した．表 4 は事前問題回答及び事後問題回答の分類結果のクロス集計表示しており，母比率の差を検定したところ $p<0.01$ で有意差が見られた．したがって，事後問題の方が，問題点と解決案両方に言及している割合が増加したことがわかる．これは，ディスカッション中で，参加者の問題についての理解が深まったためだと考えられる．

表 2 ディスカッションログ概要(男女)

参加者	ディスカッションログ数	文数
男性	79	130
女性	46	77
計	125	207

表 3 ディスカッションログ概要(文理)

参加者	ディスカッションログ数	文数
文系	67	108
理系	58	99
計	125	207

表 4　事前事後問題分類のクロス集計表

		回答の属性		合計
		問題点のみに言及している	問題点及び解決案に言及している	
事前問題回答	度数	9	5	14
	回答割合	64%	36%	100%
事後問題回答	度数	2	12	14
	回答割合	14%	86%	100%
合計	度数	11	17	28
	回答割合	39%	61%	100%

5.2. ディスカッションログの分析

ディスカッションログを読点及び改行で句切った「文」を対象とし，文献[3]の定義を参考にし，表 5 発話タイプの分類を行った．こちらの分類も著者と協力者の二人で行い，k=0.73 で分類の信頼性を確認できた．

「意見」に分類された語について，分析ツールである khcoder[11]を用い，形態素解析を行い，名詞，動詞，形容詞，形容動詞を抽出した．抽出した語彙から，階層的クラスタ分析(表 6)を行った．階層的クラスタ分析とは，語彙の共起性からクラスタを生成し，似通った文脈で使われる単語を分析する手法である．

本研究では文献[12]に基づき，特定のクラスタに属する言葉の出現率から，参加者の属性による，ディスカッション内容の違いを分析した．今回は仮説に基づき，クラスタ 1，クラスタ 2 及びクラスタ 6 について分析した．ディスカッション内容を確認したところ，クラスタ 1 では，アルゴリズムのブラックボックス化による説明責任の問題についての意見が，クラスタ 2 では，学習データから生じる性別などの不公平性についての意見が，クラスタ 6 では性別が判断基準になっていることに対する問題についての意見がそれぞれ述べられていた．

5.3. 各クラスタの語彙の出現率の差についての分析

文系と理系において，各クラスタに属する単語の文ごとの出現率についてクロス集計表を作成し，χ二乗分析を行った（表 7）．クラスタ 2 について有意差が見られ，クラスタ 2 で議論されていた，データの代表性やバイアスについての議論は理系の方が活発化していることが分かった．

男性と女性も同様に，クロス集計表を作成しχ二乗分析を行った(表 8)．すべてのクラスタで男女間での有意差は見られず，男性女性でディスカッション内容が異なるとは言えなかった．

表 5　発話タイプ一覧

タイプ番号	発話タイプ名	定義
1	意見	ケースに対する自分の意見
2	同意	相手の意見に対する同意の意見
3	範囲	相手の意見に対して反意の意見
4	反応	相手に意見を求める発言
5	その他	上記タグに分類されない発話

表 6　階層的クラスタ分析結果

クラスタ	語彙
クラスタ 1	アルゴリズム　意見　責任　確か　導入　ブラックボックス
クラスタ 2	作る　バイアス　データ　学習
クラスタ 3	必要　使う　企業　採用　人材　考える
クラスタ 4	生まれる　差　場合　特に
クラスタ 5	男性　開発　今回　感じる
クラスタ 6	性別　基準　言う　判断　倫理　問題解決　差別
クラスタ 7	可能　その　審査　説明　公平　納得提出　ES　思う　AI　人

表 7 クラスタごとの語彙出現率(文理)
（単位：文）

		理系	文系	合計	χ2 乗値
クラスタ 1	度数	10	20	30	3.016
	クラスタ割合	17%	32%	24%	
クラスタ 2	度数	13	4	17	4.836*
	クラスタ割合	22%	6%	14%	
クラスタ 6	度数	28	24	52	0.607
	クラスタ割合	47%	38%	42%	

(*:p<0.05,**:p<0.01)

表 8 クラスタごとの語彙出現率(男女)
（単位:文）

		男性	女性	合計	χ2 乗値
クラスタ 1	度数	14	16	30	1.434
	クラスタ割合	19.72%	30.77%	24.39%	
クラスタ 2	度数	9	8	17	0.027
	クラスタ割合	12.68%	15.38%	13.82%	
クラスタ 6	度数	35	17	52	2.745
	クラスタ割合	49.30%	32.69%	42.28%	

(*:p<0.05,**:p<0.01)

6. 考察

今回ケースメソッドの前後での，同じ質問に対する参加者の回答が，ケースの問題点のみへの言及から，問題点と解決策両方の言及に大きく変化した．このことは，ケースのディスカッションの中で問題に対する理解が深まり，より質の高い回答をできるようになったためだと考えられる．したがって，仮説 1 は立証され，ケースメソッドの AI 倫理への適用は有効であると考えられる．

また，ディスカッション内容の分析において，データの代表性やバイアスについての議論が行われたクラスタの単語ついて，理系の発言率が高くなった．これにより仮説 2 が立証された．すなわち，より幅広い議論をするためには，理系に対して，違った観点を与える必要がある．

一方，仮説 3 及び仮説 4 は棄却された．仮説 3 に関して，有意差は見られなかったものの，クラスタ 1 の語彙の出現率は文系が理系の倍近くになっており，データ量次第では有意差が出るのではとも考えられる．

仮説 4 に関しては，文献[7]は中学校教員のジェンダー観について分析しており，中学校教員というジェンダー問題に触れやすい仕事をしているため，学生よりもジェンダー観についての男女差があったのではないか．また，研究自体が 2000 年に行われており，時代的な変化も考えられる．

分析に関して，今回は恣意性を排除するために，階層的クラスタ分析を行い，そのクラスタに含まれる語彙の出現率でディスカッションの傾向の分析を試みたが，単に単語の出現率だけではディスカッション内容の流れをつかめないのではという批判もある．よって，こちらに関してもディスカッション内容によりタグ付けを行い，定性的分析が必要になるだろう．

また 5.2 節での文のタグ付けにおいて，タグ付け後の，各発話パターンの出現率について，文系・理系間で有意差は見られなかったものの，男性女性間では有意差が見られた．表 9 に分析結果を示す．分析の結果，その他のカテゴリに分類された文の出現率が男性の方が女性よりも高いことが分かった．特に表 1 における文系男性・理系男性のペアの実験において，19 文が発話タイプ 5 に分類された．このペアは議論に進行についての発言が多く，今回の分析の上では分析対象とならないため、今後発話タイプ 5 に分類される文が少なくなるような実験計画が必要になる．

表 9 発話タイプ一覧(男女)(単位：文)

		発話パターン					
		1	2	3	4	5	合計
男性	度数	71	13	0	15	24	123
	男性割合	58%	10%	0%	12%	20%	100.0%
女性	度数	52	5	1	16	3	77
	女性割合	68%	7%	1%	20%	4%	100.0%
	χ二乗値	1.92	0.96	1.61	2.66	9.89*	

(*:p<0.05, **:p<0.01)

7. おわりに

本研究はケースメソッドによる AI 倫理教育により，AI 倫理への理解度を自由記述式の問題から評価し，ケースメソッドの前は多くの参加者が AI についての問題点のみを記述していたが，ケースメソッドの後には問題点と解決策両方を記述し，理解度が向上したことを明らかにした．

またディスカッションをログの分析から，理系はデータの代表性やバイアスについての意見を多く発言することがわかり，理系に対して異なる観点を与えることが議論の拡大につながる．

今後の課題として，まず分析対象データの少なさが挙げられる．ケースメソッドは複数人で行うため参加者同士の組み合わせにより議論が変化する可能性があるが，各パターンの実験回数の少なさにより断念した．また，本研究ではテキストデータのみを収集しているが，実験中にアンケートなどを実施し，定量的なデータを収集することでより分析が深められる可能性がある．また本研究では文献[12]に基づきデータディスカッションログを分析したが，文献[12]は自由記述式のアンケート回答に対して分析をしており，分析している文章に連続性はない．しかし本研究では，連続性のあるディスカッションログを改行や読点で句切った文を分析しているため，より適した分析方法を適用する必要があると考えられる．

謝辞

本研究は JSPS 科研費 20K03282 の助成を受けたものです．

参考文献

[1] 数理・データサイエンス・AI（リテラシーレベル）モデルカリキュラム，入手先(PC)〈http://www.mi.u-tokyo.ac.jp/consortium/model_literacy.html〉(参照 2021-06-15)

[2] 佐藤 頌太：AI リテラシーを養う授業実践の開発 ―中学生が機械学習を用いた課題解決を行う授業実践を通じて―，千葉大学大学院人文公共学府研究プロジェクト報告書，pp.1-20(2019).

[3] 照井 賢治，菱山 玲子：多言語コミュニケーション環境における異文化分析 ヒューマンインタフェース学会論文誌 Vol.16, No.1, pp.63-76(2014).

[4] CASE CENTER JAPAN，入手先(PC)〈https://casecenter.jp/about/education.html〉(参照 2021-06-15).

[5] 島 義史：農業者育成におけるケースメソッドの現状と新規参入者への適用課題，農業経営研究第 52 巻第 1・2 号，pp.37-42(2014).

[6] 岡本 紗知：学生は理系・文系をどのように定義するのか−理論的立場に基づく考察− ，日本科学教育学会 科学教育研究，Vol44 , No.3, pp.198-207 (2020).

[7] 多々納 道子，田原 泰子：中学校教員のジェンダー観の形成要因 島根大 学教育臨床総合研究 1，pp.101-115(2001).

[8] amazon.com, 入手先(PC)〈https://www.amazon.com/〉(参照 2021-01-20).

[9] BUSINESS INSIDER：アマゾンの採用 AI ツール，女性差別でシャットダウン，入手先〈https://www.businessinsider.jp/post-177193〉(参照 2021-01-20).

[10] Richard Landis, J. and Gary G.: The Measurement of Observer Agreement for Categorical Data, Biometrics, Vol. 33, No. 1, pp. 159-174(1977).

[11] 樋口 耕一：社会調査のための計量テキスト分析，ナカニシヤ出版 (2020).

[12] 川端 亮，樋口 耕一，インターネットに対する人々の意識―自由回答の分析から―，大阪大学大学院人間科学研究科紀要第 29 巻，pp.163~181(2003).