

# データサイエンティストのための 情報倫理教育プログラムの提案

大崎理乃  
信州大学

## A Case Study on Information Ethics Education for Data Scientists

Ayano Ohsaki  
Shinshu University

As opportunities to study AI and data science continue to expand, there is a growing need to incorporate ethical education into these programs. This study proposes an educational program focused on data ethics for a specialized data science course and analyzes changes in learners' attitudes toward data ethics resulting from the program. The program, which lasted for seven weeks and covered topics such as responsible data collection and the ethical use of AI, was delivered through a combination of group discussions and project-based learning. The results suggest that the learners' interests expanded from ethical data usage to social issues.

キーワード:データサイエンス教育, Epistemic Network Analysis, 高等教育, 情報倫理, AI 倫理

### 1. はじめに

世界的にデータサイエンティスト育成のための教育プログラムの設置が多くの大学で進んでおり、その教育プログラムの検討が進んでいる<sup>(1,2)</sup>。中でも倫理教育は、社会において人工知能(AI)の利用機会が拡大するにともない、データを扱うことに関する倫理的知識・配慮が求められることから、一層の充実が求められている。

そこで、本研究では、データサイエンティスト育成のための専門課程におけるデータ利用と倫理に関する教育プログラムの開発を目的として、参加者がデータの収集と分析を行うプロジェクト型の授業を提案する。本研究における問いは、「倫理的配慮を求められるデータサイエンスプロジェクトの実践的学習の中で、学習者はどのように学習要素を認識しているのか」である。

### 2. 研究の位置づけ

#### 2.1 専門職教育における倫理教育

専門職教育における倫理教育は、技術者を対象とした「技術者倫理」<sup>(3)</sup>などで、これまでも検討されてきた。技術者教育は、国際的な枠組みもあり、国を越えてその内容や歴史が議論されてきた<sup>(4)</sup>。日本においても、学校を越えた専門職育成のための教育プログラムの検討<sup>(5)</sup>、教科書の発行など<sup>(6,7)</sup>、様々な教育リソースが充実している。一方、新たな専門職であるデータサイエンティストの教育プログラム検討では、倫理教育の重要性は議論されている<sup>(8)</sup>ものの、その専門職の特性に合わせた個別の教育プログラムの検討はいまだ少ない。なお、村田・折戸<sup>(9)</sup>による定義「ICTの利用に関わる倫理問題・社会問題を考察の対象とする情報倫理」に基づき、本稿が対象とする分野をデータの収集と利用といった情報行動に関わる情報倫理教育として

扱う。

## 2.2 提案

本研究が提案する教育プログラムでは、変化する社会において適切にデータを利用することができるようになるために、①倫理的問題の多様性、②国内外のデータ利用・AI利用に関する倫理的議論、③データ収集と利用の体験、の3点を重視することとした。授業の最終的な目標は、「AI社会で情報・データを取り扱うことについて、データ収集側とデータ提供側の両視点を踏まえて倫理的な視点から『あるべき姿』を説明できる」とした。

授業スケジュールは、表1のとおり7週間に14回の授業を実施した。授業は全てグループワークやプロジェクト活動を中心とした構成とした。前半の第1週から第3週では、倫理に関する基礎的な知識の利用と議論の練習を目的として、各回の授業テーマに関するグループディスカッションを実施した。各回授業では、グループディスカッションの内容を深めるために、教師による基礎的な知識の共有がなされた。基礎的な知識として扱われた内容は、技術者倫理教育で利用される事故のケーススタディ<sup>(6,7)</sup>、倫理学ならびに情報倫理の視点からの議論<sup>(9,10)</sup>であった。また、哲学的視点からのデータ利用と社会の発展についての議論であるポジティブな倫理とネガティブな倫理<sup>(11)</sup>のほか、AI利用と倫理に関するテーマとしてロボットに倫理を教えることを試みるモラルマシンプロジェクト<sup>(12,13)</sup>なども扱った。

後半の第4週から第7週では、グループで「幸せ」を評価するための評価式の開発に取り組んだ。この授業テーマは、「幸せ」という計量しにくく個人の価値観が強く影響する概念の評価方法開発をプロジェクト活動のゴールとすることで、プロジェクト活動中に参加者が①データの収集と生成について議論すること、②倫理的な検討が必要なデータの扱いについて議論すること、③参加者が自分ごととしてプロジェクトに関わることの3点を期待した。グループ成果物の評価は新規性、倫理性、妥当性、データ数の4観点をそれぞれ4点満点で行った。これらの評価観点のうち、本提案授

業が特に重視する倫理性については「データ収集に関する倫理的な考察がなされている」を1点、「倫理的な検討が求められるデータが一つ含まれており、そのデータ収集に関する考察がなされている」を2点、「倫理的な検討が求められるデータが二つ含まれており、そのデータ収集に関する考察がなされている」を3点、「『3点』の基準を満たした上で、特筆すべき点がある」を4点とした。評価基準はプロジェクト活動準備の第6回授業時に参加者へ伝え、参加者がプロジェクトの初期段階から、プロジェクトでは倫理的配慮についての検討が必要であることを知っている状況とした。

表1 授業スケジュール

週	回	テーマ	形式
1	1	AIと倫理	ディスカッション
	2	データサイエンスと倫理	
2	3	世界で語られるAI社会の倫理	
	4	技術発展と倫理	
3	5	ネガティブな倫理とポジティブな倫理	
	6	プロジェクト活動の準備	
4	7	プロジェクト活動1	プロジェクト
	8	プロジェクト活動2	
5	9	プロジェクト課題の発表1	
	10	プロジェクト活動3	
6	11	プロジェクト活動4	
	12	プロジェクト活動5	
7	13	プロジェクト課題の発表2	
	14	学修の振り返りとまとめ	

## 3. 方法

### 3.1 実践

提案する教育プログラムは、武蔵野大学データサイエンス学部の専門選択科目「人工知能(AI)社会の情報倫理」の中で実践した。授業中のグループは、1グループ3名から5名で構成された。本研究では、研究協力の同意を得た24名のデータを対象に分析を行う。

### 3.2 分析

授業実践において、参加者がどのような点を意識して学習、活動していたのかを把握するため、本研究では第1週から第6週までの授業後に収集した活動・学習報告の記述を分析対象とした。活動・学習報告は授業の振り返りとして実施し、授業内容に特に関連のあるものは翌週の授業で取り上げるものとして扱った。活動・学習報告の設問は「本日の授業で行った活動、活動を通して学んだこと・考えたこと、もっと知りたいと思ったことを説明してください」であった。

また、学習の成果だけでなく、参加者が期待する学習活動・内容も分析対象とするため、第1週授業後に実施した、授業における個人目標のアンケートも分析に利用した。個人目標アンケートの設問は、「14回授業が終わった時にできるようになりたいことや知ってほしいことを、その評価基準（何ができればOKかということ）とともに教えてください」であった。活動・学習報告と個人目標アンケートへの回答を合計した文章の行数は、159行であった。

分析には、数量的エスノグラフィー（Quantitative Ethnography）のツールとして利用されている Epistemic Network Analysis（以下、ENA）<sup>(14,15)</sup> を利用した。ENA は、コーディングされた発話や記述のデータを用いて、コード間の関係を2次元上に可視化するネットワーク分析の一種である。ENA の設定条件は、Unit（分析の視点）に、Phase（フェーズ）、Lesson（授業回）と Writer（活動・学習報告および学習目標の執筆者）を、Conversation（共起の計算範囲）には Lesson と Writer を用いた。

なお、ENA のためのコードは、トピックモデルを利用して生成した。トピックモデルを利用することで、文章に含まれる単語の構成を可視化し、その文章に含まれるトピックの推定が可能となる。先行研究では、論文概要を分析対象として研究領域の特徴を分析する研究<sup>(16)</sup>や、高等教育機関における授業評価アンケートの自由記述分析<sup>(17)</sup>に利用されている。本研究では、参加者による文章の特徴を捉えたコードを生成することに有用であると考え、トピックモデルを利用することとした。分析には Mathworks の MATLAB 2023a

Text Analytics Toolbox <sup>(18)</sup> を使用した。

## 4. 結果

### 4.1 コードの作成

コードを決定するためのトピックモデルによる分析の結果、上位4トピックは図1のとおりとなった。この結果に基づき、表2のコードを作成した。第1のトピックでは、倫理と AI が特徴的な語として確認されたため、コードは「AI 倫理」とした。二つ目のトピックは、「問題」を最頻出語として、倫理や思いという言葉が確認された。さらに、人間や社会という言葉も確認されたため、コードは「社会の問題」とした。第3のトピックでは、計算とデータという用語が頻出語として確認されたため、第3のコードを「データ利用」とした。最後のトピックは、特徴的な用語として確認された「幸福」をコードとして採用した。各コードの定義と文章例は表2に示す。

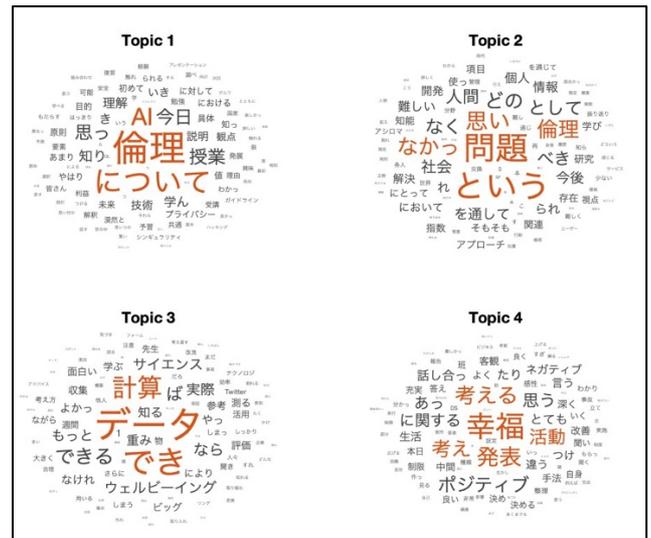


図1 トピックモデル分析の結果

コーディング後のデータを利用した ENA の分析結果を、図2から図4に示す。本研究では、授業の流れに応じた参加者の認識の変化を捉えることを目的に、授業目標アンケートへの回答を Pre、前半授業の活動・学習報告を Discussion、後半のプロジェクトの活動・学習報告を Project の三つのフェーズに分けて分析した。図中の三つの四角は、それぞれのフェーズのネットワークの中心を表し、黒丸はコードを表す。コード

表 2 コード一覧

コード	定義	例
AI 倫理	倫理および AI 倫理に関する記述	海外の倫理観や日本との倫理観の違いについての話がとても興味深かく（原文ママ）
社会の問題	データや AI 利用に関連した社会・人間活動での問題に関する記述	シンギュラリティについての議論をした
データ利用	データ利用に関する記述	アンケートの結果を分析できる形にする
幸福	幸福やウェルビーイングに関する記述	グループで幸せについての議論を行った

間をつなぐリンクは、色でフェーズを、太さでそのつながりの強さを表現する。ネットワークの色による識別は、Pre が赤、Discussion が青、Project が紫で表現されている。図 2 は授業目標アンケートの回答のみを表示させた結果であり、図 3 と図 4 は ENA の差分ネットワークの作成機能を利用して二つのフェーズの差を示した。

図 2 から図 4 までの変化は、前半の授業を通して参加者が目標に比べて「AI 倫理」と「社会の問題」を結びつけて記述していたものの、後半の授業ではあまり「AI 倫理」や「社会の問題」について記述していなかったことが示されている。具体的には、授業目標アンケートでは「AI 倫理」と「データ利用」のコードが最も強い結びつきを示しており、授業の開始時点では、参加者の認識はデータ利用と倫理に関することに向けられていたと考えられる（図 2）。続いて、Pre フェーズと Discussion フェーズの比較では、Pre フェーズを表す赤色のリンクがほとんど見えず、Discussion フェーズを表す青色のリンクは「AI 倫理」と「社会の問題」の間で最も濃くなっている（図 3）。これは、授業目標アンケートの回答に対して、Discussion フェーズの活動・学習報告で倫理と社会問題の関連が強くなっていることが示されている。最後に図 4 からは、授業スケジュールの前半と後半では、参加者の記述の特徴が大きく異なっていたことが確認できる。具体的には、Discussion フェーズでは活動・学習報告において「AI 倫理」と「社会の問題」のコード間のつながりに特徴があるのに対して、Project フェーズでは活動・学習報告において「データ利用」と「幸福」のコード間に特

徴があることが確認された。

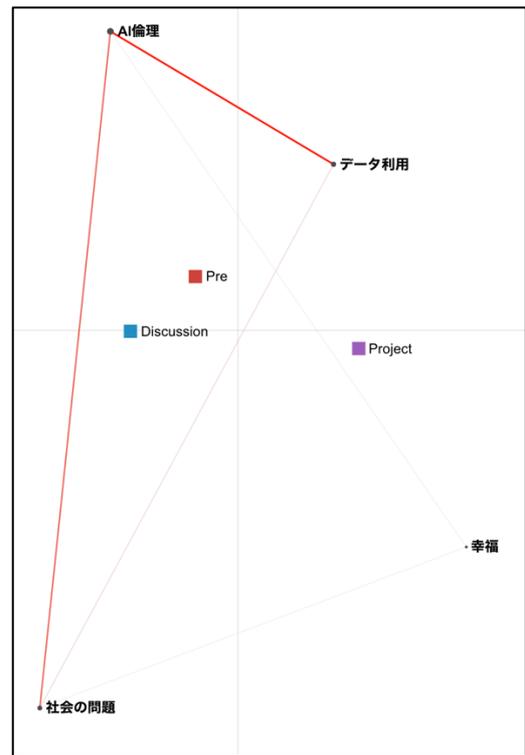


図 2 ENA による分析結果 (Pre フェーズ)

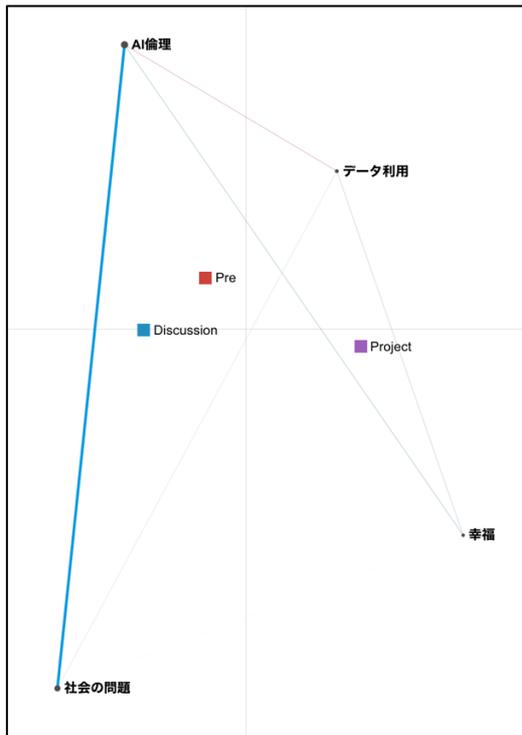


図 3 ENA による分析結果  
(Pre フェーズと Discussion フェーズの比較)

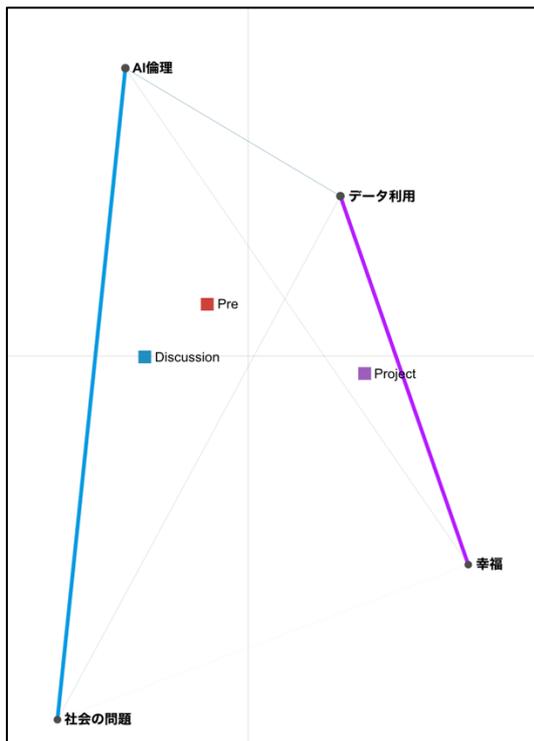


図 4 ENA による分析結果  
(Discussion フェーズと Project フェーズの比較)

## 5. まとめ

本研究では、データサイエンティストのための倫理教育プログラムを提案・実施し、その参加者の認識の変化を数量的エスノグラフィーのツールとして知られる ENA を利用して分析した。データサイエンティストの教育は、今後ますます機会が増加することが想定される。そのため、専門教育としての倫理教育プログラムの開発を目的として、プロジェクト型の教育プログラムを提案・実施した。授業は、多角的視点から継続的にデータの適切な利用方法について検討することができるようになることを重視し、知識獲得を主目的とせず、実社会に関わりのある複雑な問題に関する議論やプロジェクト活動を中心に設計した。

本研究の問いは「倫理的配慮を求められるデータサイエンスプロジェクトの実践的学習の中で、学習者はどのように学習要素を認識しているのか」であった。トピックモデルを利用したコード生成と、コーディングしたデータの ENA による可視化の結果、全 7 週間の授業スケジュールのうち、前半のディスカッション中心のフェーズでは倫理と社会の問題のつながりについて、参加者が意識していたことが示唆された。これは、各回の授業で、倫理的検討の必要となる社会問題をディスカッションテーマとして提供していたことが大きな影響と考えられる。一方で、プロジェクト活動のフェーズでは、データ収集に関する倫理的検討が参加者に求められていたにも関わらず、プロジェクトのテーマである幸せの評価式開発が活動や学習の中心になっていたことが示唆された。一般的に、プロジェクト型学習においては、学習目標よりも活動目標に参加者の意識が向きやすい。今回の実践においても同様の状況が発生していたと考えられる。なお、授業目標とディスカッションのフェーズの比較からは、授業が本格的に始まる前の参加者の興味はデータ利用と倫理の関係が中心であったものの、授業スケジュール前半では社会における倫理的問題に対象が拡張したことが示唆されている。このことは、専門学科においてデータの利用方法や分析方法を主に学んできた学生の認識を拡張させることに、本授業が貢献した可能性を示して

いると考えられる。

本研究では授業期間中の参加者の変化を捉えるために、授業の活動・学習報告を主な分析対象とした。しかし、最終的な参加者の状況を把握するためには、今後、最終レポートなどの最終成果物の評価が必要となる。また、本研究で取り組んだトピックモデルを利用した分析では、「について」「でき」など意味を持たない用語が検出されていることが確認された。今後、分析方法の精緻化を目指して、改善が必要とされる。

これらの課題があるものの、本研究ではデータサイエンスの専門教育を対象としてデータの収集や利用、AIを扱った情報倫理教育プログラムの提案と実践に取り組み、その効果だけでなく、課題としてプロジェクト活動期間には参加者の認識がプロジェクトテーマに偏っていることが示唆された。この問題については、参加者へプロジェクト活動の目的をより丁寧に周知するなど改善が期待される。今後、改善された授業のデータ分析などを通して、授業デザインの原則について検討する予定である。

## 謝辞

本研究の一部は JSPS 科研費 JP18K13238, 23K11357 の支援を受けたものである。

## 参考文献

- (1) 竹村彰通, 和泉志津恵, 齋藤邦彦, 姫野哲人, 松井秀俊, 伊達平和: “データサイエンス教育の滋賀大学モデル”, 統計数理, 第 66 巻, 第 1 号, pp. 63-78 (2018)
- (2) 情報処理学会: “データサイエンス・カリキュラム標準 (専門教育レベル) の公開”, [https://www.ipsj.or.jp/annai/committee/education/public\\_comment/kyoiku20210415.html](https://www.ipsj.or.jp/annai/committee/education/public_comment/kyoiku20210415.html), 2021 (2023 年 4 月 12 日確認)
- (3) 札野順: “技術者倫理教育, その必要性, 目的, 方法, 現状, 課題”, 工学教育, 第 54 巻, 第 1 号, 1\_16-1\_23 (2006)
- (4) 金光秀和: “技術者倫理の展望—その歴史的背景と今後—”, 情報知識学会誌, 第 16 巻, 第 3 号, 3\_24-3\_38

- (2006)
- (5) 小林幸人, 札野順: “「技術者倫理教育における学習・教育目標 2016」 および 「モジュール型モデル・シラバス」 解説”, 工学教育, 第 64 巻, 第 5 号, 5\_141-5\_159 (2016)
- (6) 札野順: “新しい時代の技術者倫理”, 放送大学教育振興会 (2015)
- (7) 金沢工業大学・科学技術応用倫理研究所: “本質から考え行動する科学技術者倫理”, 白桃書房, (2017)
- (8) ACM Data Science Task Force: “Computing competencies for undergraduate data science curricula”, Association for Computing Machinery: NY, 2021, <https://doi.org/10.1145/3453538>
- (9) 村田潔, 折戸洋子 (編著): “情報倫理入門”, ミネルヴァ書房 (2021)
- (10) 児玉聡: “実践・倫理学”, 勁草書房 (2020)
- (11) M・クーケルバーク (著), 直江清隆ほか (訳): “AI の倫理学”, 丸善出版社 (2020)
- (12) Moral machine, <https://www.moralmachine.net/hl/ja> (2023 年 4 月 13 日確認)
- (13) W・ウォラック, C・アレン (著), 岡本慎平, 久木田水生 (訳): “ロボットに倫理を教える モラル・マシーン”, 名古屋大学出版会 (2019)
- (14) Shaffer, D. W.: “Quantitative Ethnography”, Cathcart, Madison (2017)
- (15) Epistemic Network Analysis, <https://www.epistemicnetwork.org/> (2023 年 4 月 13 日確認)
- (16) Lemay, D. J., Baek, C., & Doleck, T.: “Comparison of learning analytics and educational data mining: A topic modeling approach”, Computers and Education: Artificial Intelligence, Vol.2, 100016 (2021)
- (17) 松河 秀哉, 大山 牧子, 根岸 千悠, 新居 佳子, 岩崎 千晶, 堀田 博史: “トピックモデルを用いた授業評価アンケートの自由記述の分析”, 日本教育工学会論文誌, 第 41 巻, 第 3 号, pp. 233-244 (2017)
- (18) Mathworks : “ Text Analytics Toolbox ”, <https://jp.mathworks.com/products/text-analytics.html> (2023 年 4 月 13 日確認)