

1. 研究タイトル

「日本における移民政治家の政治参加の実態と課題——最先端の大規模言語モデル（LLM）を活用したテキスト分析とフィールドワークによる複合的アプローチ」

2. 問題意識

日本は少子高齢化の進行に伴い、事実上の移民受け入れ国としての性格を強めつつある。しかしながら、移民背景を有する政治家（以下「移民政治家」）は未だ少数にとどまり、その実態や社会的影響力については十分な研究が行われていない。移民政治家は、自身の移民背景を必ずしも公にしないケースがあり、社会的偏見、メディア報道の歪み、SNS 上の差別的言説といった課題に直面しやすく、そもそもデータの取得や本人確認が難しいことから、学術的にもその正確な把握が困難である。

近年、少子高齢化や労働力不足を背景に、移民・外国人労働者の受け入れ拡大に関する政策議論が活発化する一方で、その議論の当事者となり得る移民政治家の発言・活動実態が十分に可視化されていない。加えて、日本社会におけるマイノリティへの理解不足や、SNS 上の急速な意見の拡散などが絡み合い、民主主義の正当性や多文化共生において重大な影響を及ぼす可能性がある。

こうした状況を打破するうえで、議会会議録や選挙公報、SNS 投稿、オンラインニュースなど多様なテキストデータの大規模分析が不可欠となる。しかし、政治学分野での従来型の NLP 手法だけでは、文脈や微妙な差別表現、移民背景に関連する潜在情報を十分に捉えきれない可能性が指摘されている。そこで、本研究では最先端の大規模言語モデル（GPT 系・日本語 BERT 拡張・mT5 など）の活用により、より高精度かつ柔軟なテキスト分析を実現し、移民政治家の実態把握と課題の抽出を試みる。

3. 研究の目的と意義

3.1 研究の目的

①移民政治家の実態把握と可視化

会議録や選挙公報、SNS、メディア記事など多様なテキストデータを対象に、大規模言語モデルを用いた高度な自然言語処理を行い、移民政治家の数、主張内容、社会的評価・偏見の構造を体系的に抽出・分析する。

マイノリティ性が隠されがちな政治家も含め、潜在的な移民背景や複数世代にわたる移民ルーツなど、これまで見落とされてきた要素を定量・定性の両面から捉える。

②政治・社会環境における障壁と克服戦略の解明

SNS 上の差別言説やメディアフレーミングが、移民政治家の政治活動や支持獲得に与える影響を、感情分析や stance detection など最先端の AI 技術を用いて可視化する。

フィールドワーク（インタビュー調査）を組み合わせ、移民政治家が直面する制度的・社会的障壁や、それを乗り越えるための戦略・パートナーシップ形成の実態を掘り下げる。

③新たな研究手法の確立と応用可能性の提示

大規模言語モデルを政治学・社会科学研究の文脈に導入し、テキスト分析の高度化を図る。

得られた方法論を移民政治家以外のマイノリティ（女性、LGBTQ+など）にも横展開し、広範なマイノリティ政治研究への応用を目指す。

3.2 研究の意義

①学術的意義：先進的な方法論と新知見

従来の移民政治家研究は事例数の少なさやデータ取得の難しさから、事例研究的・断片的な分析にとどまってきた。本研究では最先端 AI を活用することで、より大規模・高精度なテキスト分析を実現し、移民政治家の言説や社会的反応を包括的に捉える。

量的分析と質的分析を繰り返し行き来する混合研究法（Mixed Methods）をさらに発展させ、大規模言語モデルを社会科学理論に組み込む新たな手法論を提示する。

②社会的意義：多文化共生と民主主義の強化

マイノリティの声が十分に政治過程に反映されないと、政策形成や社会的合意に偏りが生じ、民主主義の正当性が損なわれる可能性がある。本研究の成果は、移民政治家が直面する課題を可視化し、より公正な制度設計や市民意識の醸成に寄与する。

移民政治家の活動を通じた多文化共生の推進は、長期的な人口動態や国際化に対応するうえで不可欠であり、社会全体としても意義が大きい。

③応用可能性：他のマイノリティ研究への展開

大規模言語モデルや生成系 AI によるテキスト分析手法は、移民政治家にとどまらず、女性政治家、LGBTQ+政治家、高齢・障がい者政治家など、多様なマイノリティ集団の研究にも転用可能である。

本研究を契機に、政治学・社会科学の研究手法が進化し、AI 技術を駆使した学際的なコラボレーションがさらに進むことが期待される。

4. 研究方法

本研究では、(1) 大規模テキストデータを対象とした最先端 AI (LLM) による定量的・計量的分析と、(2) フィールドワーク (インタビュー・現地調査) による質的データの収集・分析を往還させる複合的アプローチ (Mixed Methods) を採用する。以下に、研究の具体的手順と手法を示す。

4.1 データ収集・構築

①公的文書・議会会議録

対象範囲: 国会 (衆議院・参議院) の会議録、主要地方議会 (都道府県議会・政令指定都市議会など) の会議録

目的: 移民背景を持つ政治家と推測される議員の発言を抽出し、議案への立場や論点を確認する

具体的手法:

オンラインデータベース (e.g., 国会会議録検索システム) や地方議会サイトから会議録を取得

PDF 形式の文書に対しては OCR (Optical Character Recognition) を適用し、テキスト化 → 整理

「外国人」「移民」「多文化」「二重国籍」などのキーワードベース検索に加え、大規模言語モデルを活用した類義語検索やセマンティック検索も検討する

②選挙公報・選挙関連資料

対象範囲: 国政選挙・地方選挙の公報、政見放送テキスト、選挙ビラなど

目的: 候補者 (移民背景の可能性を含む) の政策訴求や自己紹介部分を可視化し、移民に関する主張の有無・ニュアンスを分析

具体的手法:

各自治体選挙管理委員会の公開情報や自治体アーカイブから資料を収集

大規模言語モデルを利用したテキスト解析 (後述) に供するため、OCR および前処

理を実施

③SNS・オンラインメディア報道

対象範囲: Twitter (X) や Facebook 等の SNS 投稿、Yahoo!ニュースや新聞社のオンライン記事など

目的: 移民政治家に対する世論の反応、差別的・支持的言説の動向を明らかにする

具体的手法:

Python のクローリング・スクレイピングツール (Scrapy 等) や公式 API を用いてデータを収集

SNS 利用規約や著作権等を遵守したうえで、必要に応じて期間を限定したデータセットを構築

データの重複排除やノイズ処理、利用規約上の制約確認を徹底する

④フィールドワーク対象者の選定

上記テキスト分析で特徴的・注目度の高い発言や、潜在的な移民背景を示唆する議員・関係者をリストアップ

関連する NPO/NGO (多文化共生支援団体など) や自治体担当者とも連携を図り、インタビュー先を抽出する

4.2 自然言語処理 (NLP) ・ AI 分析

①テキスト前処理・クレンジング

形態素解析: MeCab, Sudachi, Janome などから選定し、政治関連の固有名詞や移民関連の専門用語をユーザー辞書登録

ノイズ除去: 重複投稿やスパム、引用記事の重複箇所などを除外し、解析対象の「ユニーク発言」「オリジナルソース」を確保

品詞・文構造の整形: 後段での大規模言語モデル入力形式に合わせたトークナイズや正規化

②大規模言語モデル (LLM) の活用

モデル選定: GPT-4 系統や日本語 BERT 拡張モデル、mT5 など複数候補を比較検討

ドメイン適応 (Fine-tuning): 政治・議会関連コーパスで追加学習を行い、移民関連表現や婉曲的差別表現を識別しやすいモデルを構築

In-context Learning / Few-shot 学習: 事前に用意した人力アノテーションデータをモデルに提示し、stance detection や感情分析の精度を高める

③感情分析・stance detection

目的: 移民政治家に対するポジティブ・ネガティブ・ニュートラルな言説の分布や、当該政治家自身の発言傾向を分類

手法: LLM を用いた高精度感情分類、移民政策への「賛成・反対・条件付き」などの立場検出(stance)を実行

評価: Recall, Precision, F1 スコア等の評価指標とともに、誤判定の分析を行いモデルをチューニング

④トピック抽出・クラスタリング

BERTopic や Sentence-BERT 等の埋め込み技術を用いて、移民政治家に関する議会発言や SNS 投稿をクラスタリング

時系列分析 (いつ、どのような論点が浮上・沈静化したか) や議員間のテーマ類似度などを可視化し、移民政治家の政策アジェンダや世論の変遷を解析

⑤可視化・統計分析

統計ソフト R など、モデル出力 (感情スコア、トピック分布等) をグラフ・ヒートマップ・ネットワーク図などで可視化

移民政治家の発言量や支持動向 (SNS フォロワー数の推移など) との関連を計量モデル (回帰分析・因果推論等) で検証し、政治学的知見を得る

4.3 フィールドワーク・インタビュー

①インタビュー対象とアプローチ

LLM によるテキスト分析で抽出された「特徴的・注目度の高い移民政治家」あるいは「あまり可視化されていない潜在的移民背景の政治家」を中心に、半構造化インタビューを実施

議員事務所へのコンタクトや NPO を介したアプローチなど、倫理面・個人情報保護を踏まえて実施

②インタビュー項目の設計

政治活動上の障壁: 社会的偏見や制度的制約、SNS 上のネガティブキャンペーンの実態

移民背景の扱い：公開・非公開の判断基準、その心理的要因・メリット/デメリット

メディアとの関係：報道フレーミング、誤報・偏向の有無とその対応策

戦略的コミュニケーション：支持基盤の形成方法やネットワーク構築、差別言説への対処法

③データ整理と質的分析

録音・逐語録（可能な範囲で）の作成後、生成系 LLM による自動要約や質的解析ソフト（NVivo 等）を併用してコード化・テーマ分類

テキスト分析結果との対比を行い、LLM による定量的知見の背後にある具体的事例・主観的認識を深掘りする

④相互補完的フレームワーク（Mixed Methods）

定量分析 → インタビュー → 再度定量分析という循環プロセスを組み込み、最初の分析で発見したパターン・仮説をインタビューで検証し、それを再度定量的に裏付ける

社会科学理論（政治参加理論、アイデンティティ政治論、多文化共生理論等）との照合を随時行い、研究の理論的妥当性を高める

4.4 研究倫理・データ管理

①研究倫理審査（IRB）の実施

インタビューおよび SNS データの扱いに際し、被面接者の匿名性やプライバシー保護、同意取得の手順等を明確化

移民背景というセンシティブな個人情報扱うため、個人が特定されないようデータの扱い方を慎重に設計

②データセキュリティ・著作権対応

取得した会議録や SNS 投稿は規約・著作権法に反しない形で利用し、公開範囲を限定した管理サーバを用いる

AI モデルに投入する際のプライバシー、再学習データの保管・共有ルールなどを事前に整備

4.5 研究方法の期待される効果と課題

①効果・利点

大規模言語モデルによる高精度な言説分析で、膨大なテキスト群から移民政治家に関わる重要な論点や感情傾向を抽出できる

質的フィールドワークと組み合わせることで、モデルでは捉えにくい「隠れた戦略」「本人の認識・思考様式」を把握し、分析に深みを与える

②課題・リスク

LLM のモデルバイアスや、政治・社会文脈の解釈誤差に注意が必要

移民背景をオープンにしていない政治家へのアプローチは倫理的配慮が必須であり、研究がデリケートな政治問題に波及するリスクがある

大規模なデータ処理には時間・計算コストがかかり、予算や研究期間との調整が求められる