

ZW

博士論文

政策形成における生成AIと人間の関係性： 公共交通政策を事例として

Generative AI and Human Relationships in Policy Formation:
A Case Study of Public Transport Policy

氏名：永田 右京

指導教員：○○ 教授

○○大学 大学院 ○○研究科
○○専攻 博士後期課程

2026年○月○日提出

謝辞

本論文の作成にあたり、多大なるご指導とご支援を賜りました〇〇大学教授〇〇先生に、心より感謝申し上げます。先生の温かいご指導と厳しいご助言なしには、本論文を完成させることはできませんでした。

また、〇〇研究科の諸先生方には、研究全般にわたり貴重なご助言をいただきました。深く感謝申し上げます。

〇〇研究所の皆様には、研究の場を提供していただき、有意義な議論を重ねることができました。心より御礼申し上げます。

共同研究者の皆様、特に〇〇様には、多くのデータ収集や分析においてご協力いただきました。深く感謝いたします。

日々の研究生活を共にした研究室の仲間たちには、多くの刺激と励ましをいただきました。皆様との議論は、本研究の発展に不可欠でした。

最後に、私を支え続けてくれた家族に深く感謝いたします。皆様の理解と励ましがあったからこそ、本研究を完遂することができました。

なお、本研究の一部は、日本学術振興会科学研究費補助金（課題番号：〇〇〇〇〇〇〇〇〇）の助成を受けたものです。

2026年〇月〇日
永田 右京

要旨

本研究は、公共交通政策を舞台に、生成 AI と人間の協調的関係性を探求したものである。

第 1 章では、研究の背景、問題の所在、研究目的を論じた。公共交通政策における「連携・共創」の実装ギャップが、制度的欠陥だけでなく、人間の認知バイアスに起因する可能性を指摘した。

第 2 章では、Human-AI Policy、協調的ガバナンス論、認知バイアスと意思決定、ZK-SNARKs と政策評価に関する先行研究をレビューし、理論的空白を特定した。

第 3 章では、日本の公共交通政策の変遷と制度設計の現状を整理し、Japan MaaS 38 プロジェクトの実証分析を通じて実装ギャップの実態を明らかにした。

第 4 章では、協調ロボット制御モデルを用いた計算論的分析により、現状維持バイアスの閾値効果、確証バイアスの逆説的效果、狭い視野の一貫した負の影響を解明した。

第 5 章では、生成 AI を人間の「執政の創造性」を補完する「杖」として位置づけ、ZK-SNARKs の概念を援用した政策評価システムを提案した。Constitutional AI、市民討議、LLM as a Judge を組み合わせた三層アーキテクチャを設計した。

第 6 章では、認知バイアスへの対処戦略と生成 AI を組み込んだ制度設計への示唆を導出した。

第 7 章では、研究の総括と、都市計画への展開について論じた。

本研究は、生成 AI と人間の関係性を理論的・実証的に探求し、より良い政策形成のための指針を提供した。

キーワード：生成 AI、政策形成、認知バイアス、公共交通政策、ZK-SNARKs、制度設計

Abstract

This study explores the collaborative relationship between generative AI and humans, using public transport policy as an empirical context.

Chapter 1 discusses the research background, problem statement, and objectives. It points out that the implementation gap in "collaboration and co-creation" in public transport policy may be attributable not only to institutional deficiencies but also to human cognitive biases.

Chapter 2 reviews prior research on Human-AI Policy, collaborative governance theory, cognitive bias and decision-making, and ZK-SNARKs in policy evaluation, identifying theoretical gaps.

Chapter 3 examines the evolution of Japanese public transport policy and current institutional design, revealing the reality of implementation gaps through empirical analysis of 38 Japan MaaS projects.

Chapter 4 employs computational analysis using cooperative robot control models to uncover the threshold effect of status quo bias, the paradoxical effect of confirmation bias, and the consistently negative impact of narrow framing.

Chapter 5 positions generative AI as a "staff" that complements human "executive creativity," proposing a policy evaluation system utilizing ZK-SNARKs concepts. It designs a three-layer architecture combining Constitutional AI, citizen deliberation, and LLM as a Judge.

Chapter 6 derives strategies for addressing cognitive biases and implications for institutional design incorporating generative AI.

Chapter 7 summarizes the research and discusses future directions for urban planning applications.

This study provides theoretically and empirically grounded guidelines for better policy formation by exploring the relationship between generative AI and humans.

Keywords: Generative AI, Policy Formation, Cognitive Bias, Public Transport Policy, ZK-SNARKs, Institutional Design

Contents

| | |
|---|----------|
| 謝辞 | ii |
| 要旨 | iii |
| Abstract | iv |
| 発表論文リスト | xii |
| 1 序論 | 1 |
| 1.1 研究の背景 | 1 |
| 1.1.1 生成AIと人間の関係性という問い合わせ | 1 |
| 1.1.2 公共交通政策における「連携・共創」の潮流 | 1 |
| 1.2 問題の所在 | 1 |
| 1.2.1 協調的ガバナンスの実装ギャップ | 1 |
| 1.2.2 人間の認知限界と政策形成 | 1 |
| 1.2.3 生成AIの可能性と限界 | 2 |
| 1.3 研究目的と意義 | 2 |
| 1.4 論文の構成 | 2 |
| 2 先行研究のレビュー：Human-AI Policy の議論 | 3 |
| 2.1 はじめに | 3 |
| 2.2 Human-AI Policy：政策形成におけるAIと人間の関係性 | 3 |
| 2.2.1 AI政策論の展開 | 3 |
| 2.2.2 「執政の創造性」とAIの補完性 | 3 |
| 2.2.3 AIの「杖」としての位置づけ | 4 |
| 2.3 協調制御理論と社会システムへの応用 | 4 |
| 2.3.1 協調制御理論の基礎 | 4 |
| 2.3.2 社会システムへの応用可能性 | 4 |
| 2.4 協調的ガバナンス論 | 4 |
| 2.4.1 協調的ガバナンスの定義 | 4 |
| 2.4.2 公共交通における連携・共創 | 4 |
| 2.4.3 実装ギャップの指摘 | 5 |
| 2.5 認知バイアスと意思決定 | 5 |
| 2.5.1 行動経済学の基礎概念 | 5 |
| 2.5.2 政策プロセスにおける認知バイアス | 5 |
| 2.6 ZK-SNARKsと政策評価 | 5 |
| 2.6.1 ZK-SNARKsの基本概念 | 5 |
| 2.6.2 政策評価への応用可能性 | 5 |
| 2.7 小括：理論的空白の特定 | 6 |

| | |
|--|-----------|
| 3 舞台としての公共交通政策：現状と課題 | 7 |
| 3.1 はじめに | 7 |
| 3.2 日本の公共交通政策の変遷 | 7 |
| 3.2.1 5つの時期区分 | 7 |
| 3.2.2 STO フレームワークによる分析 | 7 |
| 3.3 制度設計の現状 | 8 |
| 3.3.1 官民連携 (PPP) と官官連携 (PuP) | 8 |
| 3.3.2 地域公共交通計画の義務化 | 8 |
| 3.4 実装ギャップの実証分析：Japan MaaS | 8 |
| 3.4.1 分析の枠組みとリサーチクエスチョン | 8 |
| 3.4.2 分析対象 | 9 |
| 3.4.3 分析結果 | 9 |
| 3.4.4 構造的要因 | 9 |
| 3.4.5 事例分析 | 10 |
| 3.4.6 日本版 MaaS は交通まちづくりの一類型として捉えられるか | 10 |
| 3.5 小括：公共交通政策を「実験場」として位置づける理由 | 10 |
| 4 認知バイアスの政策協調への影響：計算論的分析 | 12 |
| 4.1 研究背景と問題設定 | 12 |
| 4.2 Japan MaaS 政策分析：実装ギャップの実証 | 12 |
| 4.3 計算論的モデリング・フレームワーク | 13 |
| 4.3.1 協調制御モデルの構造 | 13 |
| 4.3.2 シンボル-概念対応表 | 13 |
| 4.3.3 運動学モデル | 13 |
| 4.4 認知バイアスの統合 | 13 |
| 4.4.1 現状維持バイアス | 13 |
| 4.4.2 確証バイアス | 14 |
| 4.4.3 狹い視野 | 14 |
| 4.5 シミュレーション実験 | 14 |
| 4.5.1 実験条件 | 14 |
| 4.5.2 評価指標 | 14 |
| 4.6 結果と考察 | 14 |
| 4.6.1 現状維持バイアスの閾値効果 | 14 |
| 4.6.2 確証バイアスの逆説的効果 | 14 |
| 4.6.3 狹い視野の一貫した負の影響 | 14 |
| 4.7 統計的サマリー | 15 |
| 4.8 小括：認知バイアスによる協調失敗のメカニズム | 15 |
| 5 生成 AI と人間の関係性：ZK-SNARKs 型政策評価システム | 16 |
| 5.1 はじめに：ウィキッド・プロブレムと政策評価の課題 | 16 |
| 5.1.1 ウィキッド・プロブレムの台頭 | 16 |
| 5.1.2 コミュニケーションにおける二重の困難 | 16 |
| 5.2 生成 AI の位置づけ | 16 |
| 5.2.1 人間の「執政の創造性」を補完する「杖」 | 16 |
| 5.2.2 AI ができないこと | 17 |
| 5.3 認知バイアスへの対話的介入 | 17 |
| 5.3.1 「悪魔の代理人」機能 | 17 |
| 5.3.2 視野拡大の支援 | 17 |
| 5.3.3 メタ認知の促進 | 17 |
| 5.4 ZK-SNARKs 概念の援用 | 17 |
| 5.4.1 ZK-SNARKs とは何か | 17 |
| 5.4.2 ZK-SNARKs が実用的な理由 | 18 |

| | | |
|----------|---|-----------|
| 5.4.3 | 秘密を守りながら専門性を証明する | 18 |
| 5.4.4 | ZK-SNARKs の 4 つの特性の政策評価への翻訳 | 18 |
| 5.5 | LLM as a Judge による実装 | 18 |
| 5.5.1 | LLM as a Judge と ZK-SNARKs の親和性 | 18 |
| 5.5.2 | LLM as a Judge の可能性 | 19 |
| 5.5.3 | LLM as a Judge の限界と課題 | 19 |
| 5.5.4 | Constitutional AI と市民討議 | 19 |
| 5.5.5 | 計画・推論分離型システム | 20 |
| 5.5.6 | 控訴プロセスによる人間介入 | 20 |
| 5.5.7 | 決定論的運用 | 20 |
| 5.6 | ZK-SNARKs 型政策評価のアーキテクチャ | 20 |
| 5.6.1 | 三層アーキテクチャ | 20 |
| 5.6.2 | 限界と課題 | 21 |
| 5.7 | 小括 | 21 |
| 6 | 制度設計への示唆 | 22 |
| 6.1 | 理論的含意 | 22 |
| 6.1.1 | 第 4 章のシミュレーション結果からの設計原則 | 22 |
| 6.1.2 | 第 5 章の ZK-SNARKs システムからの制度的含意 | 22 |
| 6.2 | 各バイアスへの対処戦略 | 22 |
| 6.2.1 | 現状維持バイアス：小さな変化の積み重ね | 22 |
| 6.2.2 | 確証バイアス：多様な視点の構造的導入 | 22 |
| 6.2.3 | 狭い視野：全体目標の可視化 | 23 |
| 6.3 | 生成 AI を組み込んだ制度設計 | 23 |
| 6.3.1 | 政治的-行政的インターフェース | 23 |
| 6.3.2 | 行政的-事業的インターフェース | 23 |
| 6.3.3 | ZK-SNARKs 型システムの制度的位置づけ | 23 |
| 6.4 | 「連携・共創」の再設計 | 23 |
| 6.4.1 | 三層制度設計 (S-T1-T2-O) の再考 | 23 |
| 6.4.2 | 生成 AI を「杖」として活用するガバナンス | 23 |
| 6.5 | 小括 | 24 |
| 7 | 結論：都市計画への展開 | 25 |
| 7.1 | 研究の総括 | 25 |
| 7.1.1 | 研究目的の達成 | 25 |
| 7.1.2 | 核となる主張 | 25 |
| 7.2 | 理論的貢献 | 25 |
| 7.2.1 | 認知バイアスの計算論的分析手法の政策科学への導入 | 25 |
| 7.2.2 | ZK-SNARKs 概念の政策評価への応用 | 26 |
| 7.2.3 | 生成 AI の「杖」としての理論的位置づけ | 26 |
| 7.3 | 実践的貢献 | 26 |
| 7.3.1 | 制度設計への提言 | 26 |
| 7.3.2 | ZK-SNARKs 型政策評価システムの設計指針 | 26 |
| 7.4 | 今後の課題：都市計画を舞台にした実証 | 26 |
| 7.4.1 | より複雑な政策領域への適用 | 26 |
| 7.4.2 | ZK-SNARKs システムの社会実装 | 26 |
| 7.4.3 | 生成 AI と人間の協調的関係性の継続的検証 | 27 |
| 7.5 | 結び | 27 |

| | |
|--|-----------|
| A 付録 | 29 |
| A.1 協調ロボット制御モデルの数式展開 | 29 |
| A.1.1 運動学モデル | 29 |
| A.1.2 協調係数の計算 | 29 |
| A.2 Japan Maas プロジェクト分析の詳細 | 29 |
| A.2.1 分析対象プロジェクト一覧 | 29 |
| A.2.2 評価指標のコーディング基準 | 29 |
| A.3 ZK-SNARKs 型政策評価システムの実装詳細 | 30 |
| A.3.1 システム構成 | 30 |
| A.3.2 技術的仕様 | 30 |

List of Figures

List of Tables

| | | |
|-----|-----------------------------|----|
| 3.1 | STO フレームワークによる政策分析 | 8 |
| 3.2 | Japan MaaS 38 プロジェクトの評価指標分析 | 9 |
| 4.1 | シンボル-概念対応表 | 13 |
| 4.2 | 認知バイアス効果の統計的サマリー | 15 |
| 5.1 | ZK-SNARKs 特性の政策評価への翻訳 | 18 |
| 6.1 | 認知バイアスと生成 AI を考慮した三層制度設計 | 24 |
| A.1 | 評価指標のコーディング基準 | 30 |
| A.2 | 技術的仕様 | 30 |

発表論文リスト

査読付き論文

1. Nagata, U. (2025). ”「日本版 MaaS」は「交通まちづくり」の一類型として捉えられるか？目標とガバナンスについての一考察”, 『日本評価学会誌』, Vol.XX, No.X, pp.XX-XX.
2. Nagata, U. (2025). ”Re-designing Collaboration and Co-creation in Regional Public Transport Policy: Integrated Approach to Cognitive Biases and Institutional Coordination”, 土木学会論文集, Vol.XX, No.X, pp.XX-XX.
3. Nagata, U. (2025). ”ZK-SNARKs 概念を援用した、ウィキッド・プロブレムに対応する政策評価の仕組み”, 『公共政策学会誌』, Vol.XX, No.X, pp.XX-XX.

学会発表

1. 永田右京 (2024). ”地域公共交通計画の実装ギャップに関する分析”, 日本公共政策学会 2024 年度大会.
2. 永田右京 (2024). ”認知バイアスが政策協調に与える影響の計算論的分析”, 土木学会第 XX 回年次学術講演会.
3. 永田右京 (2025). ”生成 AI と人間の協調的関係性の設計”, 日本行政学会 2025 年度大会.

Chapter 1

序論

1.1 研究の背景

1.1.1 生成 AI と人間の関係性という問い合わせ

近年、生成 AI (Generative AI) の急速な発展により、政策形成プロセスにおける AI 活用の可能性が広く議論されている。ChatGPT をはじめとする大規模言語モデル (Large Language Models, LLM) は、テキスト生成、要約、分析などのタスクにおいて人間と同等、あるいはそれ以上の性能を示す場面も増えている。しかし、AI が「人間を代替」するのではなく、「人間を補完」する関係性をどのように設計すべきかという問い合わせは、依然として未解決のままである。

本研究では、生成 AI を人間の「執政の創造性」¹を補完する「杖」として位置づけ、両者の協調的関係性のあり方を探求する。

1.1.2 公共交通政策における「連携・共創」の潮流

日本の公共交通政策においては、2002 年の規制緩和以降、「連携」と「共創」が重要な政策概念として位置づけられてきた。特に 2021 年の「ポストコロナ時代の地域交通の共創に関する検討会」(国土交通省)においては、交通事業者による地域活性化、異業種との協働、コミュニティ参画という三つの次元での「共創」が提唱されている。

しかし、こうした政策的意図にもかかわらず、実装段階では多くの課題が指摘されている。

1.2 問題の所在

1.2.1 協調的ガバナンスの実装ギャップ

公共交通政策における「連携・共創」は、制度的には整備されつつあるものの、実践レベルでは大きなギャップが存在する。例えば、Japan MaaS の 38 プロジェクトを分析した先行研究²によれば、92% のプロジェクトが事業指標のみを重視し、社会的影响やアクセシビリティ改善を評価指標に組み込んだのは 29% に留まる。さらに、市民参加の仕組みを設けたプロジェクトはわずか 5% であった。

この実装ギャップは、単なる制度的欠陥だけでなく、人間の認知特性に起因する可能性がある。

1.2.2 人間の認知限界と政策形成

人間の意思決定は、認知バイアス (cognitive biases) の影響を強く受けることが知られている [1]。特に政策形成プロセスでは、以下のバイアスが重要な影響を及ぼす：

- **現状維持バイアス (Status Quo Bias)**：変化への抵抗、現状の維持選好

¹価値判断、コミュニケーション、新たな規範の創造といった人間固有の能力

²第 3 章で詳述

- 確証バイアス (Confirmation Bias)：自己の信念を確認する情報の優先的選択
- 狹い視野 (Narrow Framing)：局所的最適化への固執、全体最適の見落とし

これらのバイアスは、ステークホルダー間の協調を阻害し、政策の実装ギャップを生む一因となっている可能性がある。

1.2.3 生成 AI の可能性と限界

生成 AI は、膨大な情報の処理、パターン認識、予測を行うことで、EBPM（証拠に基づく政策形成）を支援する強力なツールとなり得る。しかし、AI には本質的な限界も存在する：

1. 規範的判断・価値創造の不在：何が社会にとって「善い」のかを判断できない
2. 文脈理解の困難性：学習データの範囲外の「未知の状況」への適応限界
3. 「創造性」の源泉の欠如：人間的な自発的な搖らぎやアナログな現実世界の機微を再現できない

これらの限界を踏まえつつ、AI を「杖」として活用する関係性をどのように設計すべきかが問われている。

1.3 研究目的と意義

本研究の目的は、以下の三点である：

1. 公共交通政策における実装ギャップの要因として、人間の認知バイアスの影響を計算論的に解明する
2. 生成 AI と人間の協調的関係性を具体化するシステムとして、ZK-SNARKs 型政策評価システムの可能性を探る
3. 認知バイアスと生成 AI を考慮した制度設計への示唆を導出する

本研究の意義は、生成 AI と人間の関係性を理論的・実証的に探求し、より良い政策形成のための指針を提供することにある。

1.4 論文の構成

本論文は、以下の 7 章から構成される：

第 2 章 先行研究のレビュー：Human-AI Policy の議論を中心に

第 3 章 舞台としての公共交通政策：現状と課題

第 4 章 認知バイアスの政策協調への影響：計算論的分析

第 5 章 生成 AI と人間の関係性：ZK-SNARKs 型政策評価システム

第 6 章 制度設計への示唆

第 7 章 結論：都市計画への展開

Chapter 2

先行研究のレビュー：Human-AI Policy の議論

2.1 はじめに

本章では、本研究の理論的基盤となる先行研究をレビューする。特に、生成 AI と政策形成、人間-AI 協調の理論、協調的ガバナンス論、認知バイアスと意思決定、ZK-SNARKs と政策評価の五つの領域を中心に整理し、理論的空白を特定する。

2.2 Human-AI Policy：政策形成における AI と人間の関係性

2.2.1 AI 政策論の展開

公共政策における AI 活用に関する議論は、2010 年代後半から急速に発展してきた。初期の議論は、AI による行政サービスの効率化や自動化に焦点が置かれていたが、近年では AI と人間の関係性そのものが問い合わせの中心となっている [2]。

この転換の背景には、生成 AI (ChatGPT、Claude、Gemini 等) の登場がある。これらの技術は、従来の AI (分類・予測) とは異なり、創発的なテキスト生成能力を持つ。この能力は、政策文書の作成、選択肢の生成、市民との対話など、政策形成の核となるプロセスに直接関与しうる。

2.2.2 Human-Centered AI の理念

Shneiderman (2020) は、Human-Centered AI (HCAI) の理念として、以下の 2 軸マトリクスを提示している：

高自動化・低制御 AI が自律的に判断し、人間は結果を受け入れるのみ

高自動化・高制御 AI が提案を行い、人間が最終判断を下す

低自動化・高制御 人間が主導し、AI が補助的な役割を果たす

低自動化・低制御 人間も AI も十分に機能しない状態

本研究が着目するのは「高自動化・高制御」の領域である。この領域では、AI の計算能力と人間の規範判断力が相互に補完し合う。

2.2.3 「執政の創造性」と AI の補完性

行政学の観点からは、「執政の創造性」という概念が重要である。これは、公務員が直面する複雑な問題に対し、価値判断、コミュニケーション、新たな規範の創造を通じて対応する能力を指す。生成 AI は、以下の領域では人間を補完し得る：

- データの処理・分析
- 選択肢の生成・提示
- 文書作成の効率化
- 多様な視点の提示
- 認知バイアスの指摘（「悪魔の代理人」機能）

一方で、以下の領域では人間の役割が不可欠である：

- 規範的判断（何が「善い」か）
- 文脈に応じた柔軟な対応
- 新たな価値の創造
- 政治的なアカウンタビリティ
- 最終的な責任の所在

AIが「できないこと」を明示的に理解することは、AIの効果的な活用にとって不可欠である。AIは学習データの範囲内での推論しか行えず、規範的な価値判断や、学習データに含まれない文脈への適応は本質的に困難である。

2.2.4 AIの「杖」としての位置づけ

生成AIを「杖（Aaron's rod）」として位置づける視点は、AIが人間を代替するのではなく、人間の能力を補完・增幅する道具として活用する考え方である。聖書においてアロンの杖がモーセの役割を補完したように、AIは政策決定者の「杖」として機能する。

この視点からは、以下の設計原則が導かれる：

1. AIは人間の最終判断を前提とする
2. AIの限界を明示的に理解する
3. 人間-AI協調のプロセスを透明化する
4. AI自体のバイアスに対処する
5. 説明責任は常に人間が負う

この「杖」としての位置づけは、AIを「執政の創造性」を支援する道具として捉え直す視点を提供する。

2.3 協調制御理論と社会システムへの応用

2.3.1 協調制御理論の基礎

協調制御理論（Cooperative Control Theory）は、複数の自律エージェントが協調して共通の目標を達成するための制御手法を研究する分野である [3]。

この理論は、以下の特徴を持つ：

- 分散的な意思決定
- 局所的な情報に基づく協調
- 全体的な目標の達成

2.3.2 社会システムへの応用可能性

協調制御理論は、社会システムの分析にも応用可能である。特に、複数のステークホルダーが関与する政策ネットワークにおいて、各主体が自律的に行動しながら全体としての政策目標を達成するプロセスをモデル化できる。

本研究では、協調ロボット制御モデルを用いて、政策ネットワークにおけるステークホルダー間の協調を分析する（第4章で詳述）。

2.4 協調的ガバナンス論

2.4.1 協調的ガバナンスの定義

Ansell and Gash (2008) [4] は、協調的ガバナンスを以下のように定義している：

「一つまたは複数の公共機関が、非政府のステークホルダーを、合意形成志向で審議的な集団的意思決定プロセスに直接関与させる統治のあり方」

2.4.2 公共交通における連携・共創

日本の公共交通政策においては、「連携」と「共創」が重要な概念として位置づけられている [5]。Kato et al. (2009) は、コミュニティ参加型地域公共交通の成功条件として以下を指摘している：

1. 関係ステークホルダー間での認識と責任分担の共有
2. 各ステークホルダーが参加から利益を得られること
3. ステークホルダーを調整するキーパーソンの存在
4. ステークホルダーの努力が利用促進・価値向上につながること

2.4.3 実装ギャップの指摘

しかし、こうした理論的条件にもかかわらず、実践レベルでは多くの課題が指摘されている。Emerson et al. (2012) [6] は、協調的ガバナンスが直面する課題として以下を指摘している：

- 高い取引コスト
- 最小公約数的な解決策への収束
- 組織された利益による捕捉

2.5 認知バイアスと意思決定

2.5.1 行動経済学の基礎概念

Kahneman (2011) [1] は、人間の思考を「システム1（速い思考）」と「システム2（遅い思考）」に分類し、認知バイアスがシステム1の特性に起因することを示した。

2.5.2 政策プロセスにおける認知バイアス

政策形成において特に重要な認知バイアスとして、以下の三つを取り上げる：

現状維持バイアス (Status Quo Bias)

Samuelson and Zeckhauser (1988) [7] によって提唱された概念で、変化よりも現状を維持することを好む傾向を指す。

確証バイアス (Confirmation Bias)

既存の信念や仮説を支持する情報を優先的に探し、反証する情報を無視・軽視する傾向 [8]。

狭い視野 (Narrow Framing)

問題を孤立して考え、より広い文脈や長期的な影響を考慮しない傾向 [1]。

2.6 ZK-SNARKs と政策評価

2.6.1 ZK-SNARKs の基本概念

ZK-SNARKs (Zero-Knowledge Succinct Non-interactive Arguments of Knowledge) は、暗号技術の一種であり、秘密情報を公開することなく、その情報の正しさを証明する技術である [9]。ZK-SNARKs は以下の 4 つの特性を持つ：

Zero-Knowledge 証明を通して元の秘密情報を一切漏洩しない

Succinct 証明サイズが常に数百バイト程度と一定

Non-interactive 証明者から検証者への 1 回の送信で証明完了

Arguments of Knowledge 真の知識を所有している必要があり偽造不可能

2.6.2 政策評価への応用可能性

ZK-SNARKs の概念を政策評価に応用することで、「秘密を守りながら専門性を証明する」仕組みが実現可能になる。例えば、企業が自社の技術情報を公開せずに、政策課題への貢献可能性を証明できる。

本研究では、ZK-SNARKs の概念を援用した政策評価システムを LLM as a Judge と組み合わせて提案する（第 5 章で詳述）。

2.7 小括：理論的空白の特定

先行研究のレビューから、以下の理論的空白が明らかになった：

1. **認知バイアスと政策協調の接続**：協調的ガバナンスの失敗要因として認知バイアスに着目した研究は限定的である
2. **計算論的分析手法の欠如**：政策協調プロセスを計算論的にモデル化した研究は少ない
3. **生成 AI の「杖」としての理論化**：AI と人間の補完的関係性を理論的に位置づけた研究は不十分である
4. **ZK-SNARKs 概念の政策評価への応用**：暗号技術の概念を政策評価に応用した試みは先駆的である

本研究は、これらの空白を埋めることを目指す。

Chapter 3

舞台としての公共交通政策：現状と課題

3.1 はじめに

本章では、本研究の「舞台」となる公共交通政策の現状と課題を整理する。まず日本の公共交通政策の変遷を概観し、次に制度設計の現状を分析し、最後に Japan MaaS プロジェクトの実証分析を通じて実装ギャップの実態を明らかにする。

本章における分析モデルとして、政策過程論を採用する。政策過程は、目的を設定した上で現状を理解し、そのギャップを課題として認識した上でそれを解決するための方策を列挙し、実施する政策を決定するという一連の流れである。この中で、どのように目標を立て、またどのように目標を設定しているかが、本章における大まかな分析対象である。

本章でいう「交通まちづくり」の概念は、太田（2008）の述べるような「交通に関連する地域の課題への対応をベースにして、市民と行政が協働して進めるまちづくり」である。原田編（2015）の述べるところの定義である「まちづくりの目的に貢献する交通計画」とは異なり、「市民の参加と協働により発展的に進める活動プロセス」である。この中では、「交通計画における住民、市民の参加」「都市計画、都市づくりとの連携」の2点が重要視されている。

3.2 日本の公共交通政策の変遷

3.2.1 5つの時期区分

日本の地域公共交通政策は、2002年の規制緩和以降、大きく5つの時期に区分できる [10] :

第1期：規制緩和期（1997-2002） 市場原理の導入、路線許可制から届出制への移行

第2期：地域協議導入期（2003-2008） 地域公共交通会議制度の創設

第3期：民主党政権期（2009-2012） 地方分権の最大化

第4期：網形成計画期（2013-2017） ネットワーク再構築の重視

第5期：効率重視期（2018-現在） 生産性向上・効率化の強調

3.2.2 STO フレームワークによる分析

Van de Velde (1999) [11] が提唱した STO フレームワークは、公共交通政策を3つの意思決定階層で分析する手法である：

Table 3.1: STO フレームワークによる政策分析

| 階層 | 主要な意思決定事項 | 想定期間 | 主な責任主体 |
|----------------|----------------|-------|---------|
| S (Strategy) | 政策目標、長期ビジョン | 10年～ | 国、広域自治体 |
| T1 (上位戦術) | サービス水準目標、路線網骨格 | 1-5年 | 国、自治体 |
| T2 (下位戦術) | 具体的ダイヤ・ルート | 半年-2年 | 交通事業者 |
| O (Operations) | 日常運行、維持管理 | 日々～月 | 交通事業者 |

3.3 制度設計の現状

3.3.1 官民連携 (PPP) と官官連携 (PuP)

公共交通政策における連携形態は、大きく分けて官民連携 (Public-Private Partnership, PPP) と官官連携 (Public-Public Partnership, PuP) がある。

官民連携 (PPP) の事例

- 空港コンセッション（北海道・関西エアポート）
- 駅無人化への地域対応（滝沢市とJR東日本）
- デマンド交通の導入（雲南省だんだんタクシー）

官官連携 (PuP) の事例

- 横の連携：盛岡都市圏地域公共交通計画
- 縦の連携：地域公共交通会議制度
- 階層型連携：前橋都市圏の自治体バス広域連携

3.3.2 地域公共交通計画の義務化

2014年の法改正により、地域公共交通網形成計画の策定が自治体に義務化された。この制度は、国がSレベルの枠組みを提供し、自治体がT1レベルで具体化する役割分担を想定している。しかし、自治体の資源制約により、形式的な計画策定に留まるケースも指摘されている。

3.4 実装ギャップの実証分析：Japan MaaS

3.4.1 分析の枠組みとリサーチクエスチョン

本章では、いわゆる「日本版MaaS」を「交通まちづくり」実践の一類型として捉え、以下の2つの問い合わせについて検討する：

- 「日本版MaaS」の政策進捗において目標設定がどこに置かれているか
- 立場を問わない市民による批判回路がどのように整備されているのか

本研究は、日本版MaaSを「MaaS」としてではなく、「交通まちづくり」の一類型として分析し、その目的意識と計画における責任体系の構築状況、すなわちガバナンス状況を探るものである。分析手法について、「計画における計画指標 (Key Performance Indicator, KPI) に事業利用状況以外の指標がある」「会議体において、交通事業者、システム供給者、学識者、自治体以外の市民参加がある」の二つの点を検討し、交通政策を地域政策として運営できているかの評価とする。

3.4.2 分析対象

日本版 MaaS は、その事業概要を一定のフォーマットに落とし込んで申請する仕組みとなっており、国土交通省は採択のたびにそれらをまとめた文書を公開している。これらには協議会の構成員や目標数値が記載されており、これを分析対象とする。

2020 年の日本版 MaaS 推進事業では、全国 36 の事業が認定されており、提出された計画については全件採択となっている。本事業によって実施された事業の特性を説明するために、日本版 MaaS として展開された各事業について、要素技術の利用状況を確認し、類型化を試みる。

日本版 MaaS の主な要素技術として、サブスク・定期など企画乗車券、オンラインで交通情報が見られる、あるいは決済できるアプリケーション、病院、商業クーポンなど外部情報の提供、デマンド交通、シェアサイクル等の新モビリティ、さらにデータ連携基盤の構築や利用が認められた。特にアプリケーションの提供が 31 事例、外部情報との連携が 32 事例と多かったものの、企画乗車券の導入と新モビリティの導入比率は低く、またデータ連携基盤の利用や構築は 12 事例と少なかった。

3.4.3 分析結果

分析の結果、双方の判断を満たし、地域政策として日本版 MaaS を運用している事例は存在しなかった。

第一に、交通事業や MaaS 事業の指標を導入している事業は 32/36 であったが、それ以外の地域目標を設定している事業は 10 のみであった。第二に、自治体が参加している事業に限定すると、パブリックコメントや協議会への市民参加を確認できた事業は 2/32 にとどまった。

Table 3.2: Japan MaaS 38 プロジェクトの評価指標分析

| 指標タイプ | プロジェクト数 | 割合 |
|----------|---------|-----|
| 事業指標のみ | 32 | 89% |
| 非事業指標を含む | 10 | 28% |
| 市民参加の仕組み | 2 | 6% |

この結果は、「連携・共創」の政策的意図が、実践レベルでは事業効率性の追求に偏向していることを示している。

3.4.4 構造的要因

実装ギャップの背景には、以下の構造的要因が存在する：

1. **自治体の資源制約**：「逆三角形」負担構造（地方分権改革推進有識者会議, 2023）
2. **技術的複雑性**：MaaS 導入に必要な専門知識の不足
3. **ステークホルダー間の利害対立**：交通事業者、自治体、市民の間での優先順位の相違
4. **認知的要因**：意思決定における認知バイアスの影響（第 4 章で詳述）

新たなガバナンスシステムの導入へのコストが高いことも要因として挙げられる。一般に、デジタル時代の政府活動への移行には、それに即した評価・批判のシステムへの移行が必要であるが、それには取引費用がかかりすぎることが分かっている。また、自治体の持つ政策評価システムが業績評価に偏っている点も指摘されている。三重県の「さわやか運動」から始まった政策評価では、担当者が直接事業内容を説明する「事務事業評価」が核であり、その中で簡略化された評価として業績評価が利用されることが多い。これは一種の日本の評価システムの文化であり、日本版 MaaS もその流れを汲んでいる可能性がある。

3.4.5 事例分析

分析の射程を示すため、2つの事例を取り上げる。

地方版 MaaS の広域連携基盤構築モデル事業（ひたち圏域）

日立市を中心とする茨城県北部の圏域における MaaS 事業展開について、3市1村の協力の下で行われた日本版 MaaS 事業である。

本事業の目標は3つに分かれている。一つ目は「利用者関連指標」であり、取組ページへのアクセス数、アプリDL数、チケット販売数などが設定されている。二つ目は「交通事業者関連指標」であり、参加する事業者数により達成される。三つ目は「MaaS 事業者関連指標」であり、接続事業者数を目標と据えている。

市民参加状況については、市民団体らしき仕組みは確認できず、市民代表者の参加形跡も確認できなかった。また、市民参加、パブリックコメントなどの情報を検索したが、見当たらなかった。

鞆の浦 MaaS

広島県福山市の景勝地、鞆の浦における MaaS 事業展開である。本事業の目標は、デジタルチケット発行数、アンケート指標（満足度70%以上など）、電動レンタサイクル利用者の回遊エリア拡大状況に分かれている。

市民参加状況については、市民団体らしき仕組みは確認できなかったが、観光関連団体の代表として「公益社団法人福山観光コンベンション協会」が出席しており、これは交通事業者外の参加と認められる。

3.4.6 日本版 MaaS は交通まちづくりの一類型として捉えられるか

以上の議論を踏まえて、日本版 MaaS を交通まちづくりとして捉えられるかについて議論する。交通まちづくりの基本線は、初期構想段階では市民参加を含んでいた以外は一貫しており、地域の課題を解決する方針のもと交通を計画し、提供するものであった。

これを踏まえると、日本版 MaaS はその構想段階では交通まちづくりの要件に合うものの、実装段階においてはその要件に合致しないと理解できる。前者については、日本版 MaaS の特徴として地域課題を解決する交通サービスの提供を支援する政策である点が該当する。一方で後者については、日本版 MaaS の計画において交通サービスの利用状況が評価の中心にあり、必ずしも地域課題解決の状況をモニタリングしていない点が該当する。

3.5 小括：公共交通政策を「実験場」として位置づける理由

公共交通政策は、以下の理由から本研究の「実験場」として適している：

1. 複雑なステークホルダー関係：国、自治体、交通事業者、市民など多様な主体が関与
2. 明確な政策目標：持続可能な公共交通の維持・発展
3. 実装ギャップの可視化：制度的意図と実践の乖離が観察可能
4. データの入手可能性：政策文書、統計データへのアクセスが比較的容易

本章の分析から、以下の示唆が得られた：

第一に、日本版 MaaS は構想段階では「交通まちづくり」と捉えられるが、実装段階ではそうではない。政策的意図として「連携・共創」が謳われても、実践レベルでは事業効率性の追求に偏重してしまう構造的課題が存在する。

第二に、交通政策の理念として、交通の機能について「地域の活力に貢献」するように施策を運営するためには、少なくとも地域がどのようになるべきかの指標を入れる必要がある。交通事業や MaaS 事業の指標のみでは、こうした運営には不十分である。

第三に、行政機関が手がける交通サービスに対する批判回路についても、十分に構築されていない状況が見えてきた。公共交通政策は多くの場合、土木工学を主軸とした専門性を問われることから、こうした回路が機能しにくいと推察される。

「利用者視点」を導入しているとうたう交通政策が各地で展開されているが、少なくとも政策において相手にするのは市民であり、この国の主権者である。専門性が必ずしも正しさではなく、目指す方向性はイデオロギーであり批判されうることを前提として、民主主義国家の計画のあり方を交通の面から探る作業が求められる。

次章では、この実装ギャップの認知的要因を計算論的に分析する。

Chapter 4

認知バイアスの政策協調への影響：計算論的分析

4.1 研究背景と問題設定

日本の地域公共交通政策は、2002年の規制緩和以来、「連携」と「共創」を強調するアプローチへと大きく変容してきた。この変化は、多様なアクター—政府機関、交通事業者、市民、他産業—の専門知識とリソースを活用し、複雑な政策課題に取り組む協調的アプローチへの、より広範な公共交通ガバナンスの潮流を反映している。

2021年の国土交通省による「ポストコロナ時代の地域交通の共創に関する検討会」は、特にこの政策動向を象徴している。この枠組みは、「共創的交通」を三つの重要な次元で明示的に求めている：地域コミュニティの活性化に貢献する人の流れの生成、交通事業者と他産業間のクロスセクター協働、そして交通を共有責任として捉えるコミュニティ参画である。

しかし、協調と共創への政策的強調にもかかわらず、実証的証拠は政策意図と実装成果の間に大きなギャップがあることを示唆している。協調的枠組みの拡大は、必ずしもより効果的または持続可能な交通ソリューションに直結しておらず、公共交通ガバナンスにおいて協調と共創が有効に機能する条件について根本的な問い合わせを提起している。

4.2 Japan MaaS 政策分析：実装ギャップの実証

Japan MaaSイニシアチブの全38プロジェクトの分析は、協調的ガバナンスにおける重要な実装ギャップを明らかにした。2020年度Japan MaaSイニシアチブで認定された全プロジェクトの分析結果：

- **事業指標**：92%のプロジェクト（35/38）が、利用率、収益生成、運行効率に焦点を当てた指標を含んでいた
- **非事業指標**：わずか29%のプロジェクト（11/38）が、社会的影響、アクセシビリティ改善、環境成果に対処する指標を組み込んでいた
- **市民参加**：わずか5%のプロジェクト（2/38）が、プロジェクトガバナンスと評価における市民参加の意味あるメカニズムを確立していた

これらの発見は、協調的枠組みがより広範な公共目的に奉仕するのではなく、交通事業者の利益によって捕捉されている可能性を示しており、真の民主的参加の最小限の達成にとどまっている。

4.3 計算論的モデリング・フレームワーク

4.3.1 協調制御モデルの構造

本研究は、Yoshihara et al. (2009) の協調ロボット制御理論を応用し、政策実装におけるステークホルダー間の相互作用をモデル化する。N 関節ロボットアームは政策ネットワークを表現し、各関節 i は以下の状態変数を持つ政策ステークホルダーに対応する：

- 位置ベクトル: $\mathbf{x}_i \in \mathbb{R}^2$
- 速度ベクトル: $\mathbf{v}_i \in \mathbb{R}^2$
- 関節角度: $\theta_i \in \mathbb{R}$
- リンク長: $a_i \in \mathbb{R}^+$

このマッピングにより、特定の能力（リンク長）と政策立場（関節角度）を持つ自律エージェントとしての政策ステークホルダーを表現し、集合的な政策目標を達成するために協調しなければならない関係をモデル化できる。

4.3.2 シンボル-概念対応表

Table 4.1: シンボル-概念対応表

| 記号 | 数学的意味 | 制度的意味 | パラメータ範囲 |
|------------|------------|-----------------|----------------------|
| θ_i | 関節角度 | ステークホルダーの政策立場 | $(-\infty, +\infty)$ |
| a_i | リンク長 | ステークホルダーの影響力・能力 | \mathbb{R}^+ |
| $b_{sq,i}$ | 現状維持バイアス係数 | 変化への抵抗 | $[0, 1]$ |
| $b_{cf,i}$ | 確証バイアス係数 | 自己判断への過信 | $[-1, 1]$ |
| $b_{nf,i}$ | 狭い視野係数 | 局所最適化への焦点 | $[0, 1]$ |
| k_i | 協調係数 | 他者との協調度 | $[0, 1]$ |

4.3.3 運動学モデル

各関節の位置は、前の関節の位置と累積関節角度に依存する：

$$\mathbf{x}_i = \mathbf{x}_{i-1} + a_i \begin{bmatrix} \cos(\sum_{j=0}^{i-1} \theta_j) \\ \sin(\sum_{j=0}^{i-1} \theta_j) \end{bmatrix} \quad (4.1)$$

ここで $\mathbf{x}_0 = \mathbf{0}$ である。これは、政策成果に対するステークホルダーの決定の累積効果を表現している。

4.4 認知バイアスの統合

4.4.1 現状維持バイアス

現状維持バイアス $b_{sq,i} \in [0, 1]$ は、関節角度変化への抵抗として組み込まれる：

$$\frac{d\theta_i}{dt} = (1 - b_{sq,i}) \cdot \omega_i \quad (4.2)$$

ここで ω_i はバイアスなしの角速度である。

4.4.2 確証バイアス

確証バイアス $b_{cf,i} \in [-1, 1]$ は、協調係数の修正として組み込まれる：

$$k'_i = k_i \cdot (1 + b_{cf,i} \cdot 0.5) \quad (4.3)$$

4.4.3 狹い視野

狭い視野 $b_{nf,i} \in [0, 1]$ は、全体目標ベクトルの削減として組み込まれる：

$$\mathbf{v}'_{tilda,i} = (1 - b_{nf,i}) \cdot \mathbf{v}_{tilda,i} \quad (4.4)$$

4.5 シミュレーション実験

4.5.1 実験条件

8関節ロボットアームを用い、以下の条件下でシミュレーションを実施した：

- 制御ゲイン $V = 1.0$
- 目標速度ゲイン $G_t = 0.5$
- シミュレーション時間：30秒
- 各バイアス値：0.0～1.0の範囲で段階的に変化

4.5.2 評価指標

以下の評価指標を用いた：

精度 × 距離 (A_t) 政策目標への到達精度と移動距離の積

エネルギー効率 (E_t) 関節動作の効率性

関節活動度 ($J_{act,i}$) ステークホルダーの活動レベル

4.6 結果と考察

4.6.1 現状維持バイアスの閾値効果

現状維持バイアスには、顕著な閾値効果が観察された。バイアス値が約0.25を超えると、協調効率が急激に低下する。

この閾値は、bounded confidence モデルにおける合意形成の閾値 $\varepsilon = 0.5$ の半分に相当し、政策変化を受け入れる「臨界点」として解釈できる。

4.6.2 確証バイアスの逆説的效果

興味深いことに、確証バイアスは適度な範囲 ($b_{cf} \approx 0.3$) では協調を促進する効果が観察された。これは、一定の自信が意思決定の迅速化に寄与するためと解釈できる。

しかし、 $b_{cf} > 0.5$ では協調が阻害され、他者からのフィードバックが無視される傾向が強まった。

4.6.3 狹い視野の一貫した負の影響

狭い視野は、一貫して協調効率を低下させた。特に、複数のステークホルダーが同時に狭い視野を持つ場合、全体最適から大きく乖離した局所解に収束する傾向が観察された。

4.7 統計的サマリー

Table 4.2: 認知バイアス効果の統計的サマリー

| バイアス種別 | 効果の方向 | 閾値 | 推奨対処 |
|----------|---------|--------------------|----------|
| 現状維持バイアス | 非線形閾値効果 | ≈ 0.25 | 段階的導入 |
| 確証バイアス | 逆説的促進 | ≈ 0.3 (最適) | 適度な自信の許容 |
| 狭い視野 | 一貫して負 | なし | 全体目標の可視化 |

4.8 小括：認知バイアスによる協調失敗のメカニズム

シミュレーション実験から、以下の知見が得られた：

1. 現状維持バイアスには閾値効果があり、臨界点を超えると急激に協調が阻害される
2. 確証バイアスは適度な範囲では逆説的に協調を促進する可能性がある
3. 狹い視野は一貫して負の影響を持ち、全体最適を損なう

これらの知見は、第 6 章での制度設計において重要な示唆となる。

Chapter 5

生成AIと人間の関係性：ZK-SNARKs 型政策評価システム

5.1 はじめに：ウィキッド・プロブレムと政策評価の課題

5.1.1 ウィキッド・プロブレムの台頭

現代社会においては、単一の専門知識や価値観では解決困難な複雑な政策課題、いわゆるウィキッド・プロブレムが頻発している。これらの問題に対処するためには、多様な関係者がそれぞれの専門知識や経験を持ち寄り、協力して解決策を模索する必要がある。しかし、関係者間の価値観の相違や情報の非対称性、プライバシーへの懸念などが、効果的な協働を阻害する要因となっている。従来の政策評価においては、専門家によるトップダウン的な評価や、ロジック・モデルを採用したカスケード式の目標設定・評価が主流であった。しかし、ウィキッド・プロブレムにおいては、その複雑さから、専門家のみによる評価では十分な解決策が得られない場合がある（Head, 2022）。また、関係者間の対話を通じたボトムアップ的な解決策の模索も試みられているが、ここでも情報の非対称性や信頼性、プライバシー保護の観点から課題が残る。

5.1.2 コミュニケーションにおける二重の困難

こうした問題点は、コミュニケーションにおいて正確な理解はそもそも非常に難しい、また知識の非対称性が「正しい情報」への理解を困難にしている、という二重の状況から生じている。前者に関しては、ルーマンの述べるコミュニケーションの3段階である情報、伝達、理解のうち、理解の多様性は非常に大きく、このエントロピーを削減する作業が非常に困難である。社会を構成するコミュニケーションの連鎖を考慮に入れると、そこに正確性を組み込むのは非常に難しい。後者に関しては前者と関連する議論として、今まで連鎖してきたコミュニケーションとしての前提知識が合致しなければ、専門家ないし「ある人」の価値観を正確には共有できない問題がある。だからこそ、コミュニケーションの背景と切り離し、一方で共有できる価値観をベースにした評価制度こそが求められる。

5.2 生成AIの位置づけ

5.2.1 人間の「執政の創造性」を補完する「杖」

第2章で述べた通り、生成AIは人間の「執政の創造性」を補完する「杖」として位置づけられる。この関係性は、以下の原則に基づく：

1. AIは人間の最終判断を前提とする
2. AIの限界を明示的に理解する
3. 人間-AI協調のプロセスを透明化する

4. AI 自体のバイアスに対する対処

5.2.2 AI ができないこと

生成 AI には、以下の本質的な限界が存在する：

規範判断 何が社会にとって「善い」のかを判断できない

価値創造 新たな価値や規範を創造できない

文脈理解 学習データの範囲外の状況に適応できない

これらの領域は、人間の役割として残される。

5.3 認知バイアスへの対話的介入

5.3.1 「悪魔の代理人」機能

生成 AI は、意思決定プロセスにおける「悪魔の代理人（Devil's Advocate）」として機能し得る。具体的には：

- 意思決定者の視点と対立する論点の提示
- 見落とされがちなリスクの指摘
- 代替案の生成

これは、確証バイアスへの対処に特に有効である。

5.3.2 視野拡大の支援

狭い視野への対処として、AI は以下の機能を提供できる：

- 全体的な目標との整合性の確認
- 他領域との関連性の提示
- 長期的な影響の分析

5.3.3 メタ認知の促進

AI は、意思決定者自身の認知バイアスへの気づきを促すことができる。例えば、「あなたの判断は現状維持バイアスの影響を受けている可能性があります」といったフィードバックを提供する。

5.4 ZK-SNARKs 概念の援用

5.4.1 ZK-SNARKs とは何か

ZK-SNARKs (Zero-Knowledge Succinct Non-interactive Arguments of Knowledge) は、証明者が検証者に対して秘密情報を開示することなく、その情報の正しさを証明する技術である。この技術は 4 つの特性から成り立つ：

Zero-Knowledge (ゼロ知識) 証明を通して元の秘密情報が一切漏洩しない

Succinct (簡潔) どんなに複雑な計算であっても証明サイズが常に数百バイト程度と一定

Non-interactive (非対話) 証明者から検証者への 1 回の送信で証明が完了し、複数回のやり取りを必要としない

Arguments of Knowledge (知識の論証) 証明者が本当にその知識を所有している必要があり、偽造が数学的に不可能

5.4.2 ZK-SNARKs が実用的な理由

ZK-SNARKs が実用的な理由は、従来のゼロ知識証明システムが持っていた複数の根本的な限界を同時に解決したことにある：

1. **処理速度の革新**：検証時間が数ミリ秒で完了し、従来の数時間から数日という検証時間を大幅に短縮
2. **証明サイズの一定性**：100 万ステップの複雑な計算であっても証明サイズは 288 バイト程度と一定
3. **適用範囲の広さ**：あらゆるコンピュータ処理を証明対象とすることが可能
4. **信頼性の高さ**：楕円曲線暗号の安全性に基づいて数学的な偽造が不可能

5.4.3 秘密を守りながら専門性を証明する

ZK-SNARKs の概念を政策評価に応用することで、「秘密情報を公開することなく、その情報が正しいことを証明する」仕組みが実現可能になる。

これは、以下の政策場面で有用である：

- 企業が技術の詳細を公開せずに、政策課題への貢献可能性を証明
- 個人が個人情報を守りながら、専門性を証明
- 行政が内部情報を守りながら、政策判断の根拠を説明

5.4.4 ZK-SNARKs の 4 つの特性の政策評価への翻訳

Table 5.1: ZK-SNARKs 特性の政策評価への翻訳

| ZK-SNARKs 特性 | 政策評価への翻訳 |
|------------------------|---------------------|
| Zero-Knowledge | 秘密情報（企業秘密・個人情報）の非開示 |
| Succinct | 簡潔な評価結果の提示 |
| Non-interactive | 一方向の対話での評価完結 |
| Arguments of Knowledge | 専門知識に基づく証明 |

5.5 LLM as a Judge による実装

5.5.1 LLM as a Judge と ZK-SNARKs の親和性

ZK-SNARKs 型システムにおける LLM as a Judge（大規模言語モデル判定者）の導入は、評価プロセスの自動化と一貫性向上において大きな可能性を秘めている。

LLM as a Judge が ZK-SNARKs フレームワークと本質的に合致する理由は、両者が「複雑な判断プロセスを单一のアルゴリズム処理に集約する」という共通の設計思想を持つことにある。ZK-SNARKs が任意の計算を多項式表現に変換して单一の証明アルゴリズムで処理するように、LLM as a Judge は多様で複雑な評価基準を統一的な言語モデル処理に収束させる。この収束性により、主観的で曖昧な人間判断を客観的で再現可能なアルゴリズム判断に置き換えることが可能となり、ZK-SNARKs の「アルゴリズムによる確からしさの提供」という目標と完全に一致する。

5.5.2 LLM as a Judge の可能性

LLM as a Judge の可能性として、以下が挙げられる：

スケーラブルな自動評価 2024-2025 年の研究により、LLM は「人間のような推論と意思決定プロセスを模倣」し、従来専門家に依存していた役割の「費用効率的でスケーラブルな代替手段」を提供することが確認されている。多基準意思決定（MCDM）フレームワーク（AHP、TOPSIS、VIKOR 等）との統合により、複雑な政策評価基準を体系的に処理することが可能となる。

一貫性のある判定 ペアワイズ比較手法により複数の政策提案を一貫した基準で比較評価し、「陪審員システム」（複数 LLM によるアンサンブル評価）を通じて判定信頼性を向上させることができる。

秘匿性と効率性の両立 評価対象の詳細を開示せずに判定結果のみを提供する構造が ZK-SNARKs の秘匿性要件と適合し、大量の提案や複雑な評価基準への対応が人的リソースの制約を解決する。

5.5.3 LLM as a Judge の限界と課題

しかし、LLM as a Judge には重要な限界と課題が存在する：

多様なバイアス問題 評価順序による判定の偏りであるポジションバイアス、LLM が自身の生成内容を過大評価する自己強化バイアス、権威ある情報源からの内容を根拠に関わらず高評価する権威バイアスなどが確認されている。

人間判断との不整合 2025 年時点でも最先端判定 LLM（GPT-4 シリーズ等）の人間との一致率は 0.7 以下であり、多言語環境では一貫性がさらに低下する。

信頼性と透明性の問題 詳細なスコアリングにおける恣意性の増大、敵対的攻撃に対する脆弱性、評価プロセスの説明可能性不足が指摘されている。

これらの課題は、政策評価における公正性と透明性の確保を困難にする要因となる。

5.5.4 Constitutional AI と市民討議

LLM as a Judge システムにおいて人間が結果を意図的に歪めるのではなく、アルゴリズム自体の改善に寄与する複数の支援機能が確立されている。その基盤となるのが、Anthropic (2022) により提案された Constitutional AI 手法である。

この革新的な手法は、人間のラベル付けを最小限に抑えながら AI の行動を制御することを可能にした。この手法は 3 つの核心原則に基づく：

1. **憲法的ルールセット**：人間が事前に定義した行動規範・価値基準のリスト
2. **自己批判・自己修正機能**：AI が自身の回答を憲法に照らして評価し改善
3. **最小限の人間監督**：詳細な判定ではなく原則設計のみに人間が関与

実装プロセスは 2 段階から構成される。教師あり学習段階では初期モデルからサンプルを生成し、AI が自己批判と修正を行って修正された応答で元のモデルを微調整する。強化学習段階では、微調整されたモデルからサンプルを生成し、別の AI モデルがサンプルを評価して選好モデルを訓練し、この選好モデルを報酬信号として RL from AI Feedback (RLAIF) を実施する。

この仕組みにより、人間は直接的な判定者ではなく、システムの設計者・調整者として機能し、複雑な価値判断を明確なアルゴリズム処理に変換することで、主観的判断の客観化が実現される。

LLM as a Judge において、評価基準は Constitutional AI の原則と市民討議を通じて設計される。これにより：

- AI の評価基準に人間の価値観を組み込む
- 民主的正当性を確保する
- 透明性と説明責任を満たす

5.5.5 計画・推論分離型システム

EvalPlanner（2025 年）に代表される最新手法では、計画コンポーネントと推論コンポーネントを分離し、人間が評価手順の設計に関与しながら LLM が推論実行を担当する役割分担が確立されている。この分離により、人間が LLM の論理プロセスを読解・修正することが容易になり、意図しない結果が生じた際のアルゴリズム改善が可能となる。

5.5.6 控訴プロセスによる人間介入

控訴プロセス型の人間介入システムでは、AI が初期判定を行い、問題のあるケースのみ人間専門家が再検討する段階的介入が実現されている。このアプローチは結果の意図的歪曲を避けながら、システム改善を図る効果的な手法として評価されている。

最終的な判断は人間が行うため、控訴プロセスを組み込む：

1. AI による一次評価
2. 評価結果に対する異議申立ての受付
3. 人間による二次評価
4. 最終判断の提示

加えてハイブリッド評価システムとして複数 LLM によるアンサンブル評価と人間による検証を組み合わせ、メタジャッジフレームワークにより判定信頼性を向上させる手法も確立されている。これらの研究成果は、LLM as a Judge システムにおいて人間がアルゴリズムの改善者として機能し、結果の歪曲者ではなくシステムの設計者・調整者として関与する新しいパラダイムを示している。

5.5.7 決定論的運用

LLM の確率的な出力を制御するため、以下の手法を組み合わせる：

- Temperature=0：ランダム性を排除
- Self-Consistency：複数回の出力から一貫性のある結果を選択
- TEE（秘匿実行環境）：処理の透明性を確保

5.6 ZK-SNARKs 型政策評価のアーキテクチャ

5.6.1 三層アーキテクチャ

本研究が提案する ZK-SNARKs 型政策評価システムは、以下の三層アーキテクチャから構成される：

外層：ZK-SNARKs 秘匿証明層 秘密情報の保護、秘匿化処理

中層：LLM as a Judge 評価層 自動評価、一貫性確保

内層：Constitutional AI + 市民討議 価値統合、評価基準の設計

5.6.2 限界と課題

このシステムには以下の限界がある：

1. **数学的保証と確率的期待の違い**：ZK-SNARKs の数学的完全性は、LLM では実現できない
2. **AI 自体のバイアス**：学習データに含まれるバイアスが評価結果に影響
3. **透明性の限界**：LLM の内部処理の完全な説明は困難

これらの限界に対処するため、人間による最終判断を不可欠とする。

5.7 小括

本章では、生成 AI と人間の協調的関係性を具体化するシステムとして、ZK-SNARKs 型政策評価システムを提案した。このシステムは：

1. 認知バイアスへの対話的介入を通じて、より良い意思決定を支援
2. 秘匿性と信頼性を両立する評価プロセスを提供
3. 人間による最終判断を前提とした、AI の「杖」としての活用を実現

次章では、第 4 章の計算論的分析と本章のシステム設計を踏まえ、制度設計への示唆を導出する。

Chapter 6

制度設計への示唆

6.1 理論的含意

6.1.1 第4章のシミュレーション結果からの設計原則

第4章の計算論的分析から、以下の設計原則が導かれる：

現状維持バイアスへの対処 臨界点 ($b_{sq} \approx 0.25$) を超えないよう、段階的な変化導入を設計する

確証バイアスの活用 適度な自信は協調を促進するため、完全な中立性よりも構造的な多様性を確保する

狭い視野への対処 全体目標の可視化と横断的指標の導入を必須とする

6.1.2 第5章のZK-SNARKsシステムからの制度的含意

ZK-SNARKs型政策評価システムは、以下の制度的含意を持つ：

1. **秘匿性の制度化**：企業秘密や個人情報を保護しながら政策参加を可能にする制度
2. **自動評価の境界**：AIによる評価と人間による判断の境界を明確化する
3. **市民参加の質的向上**：市民討議を通じた評価基準の共同設計

6.2 各バイアスへの対処戦略

6.2.1 現状維持バイアス：小さな変化の積み重ね

現状維持バイアスへの対処として、以下の戦略を提案する：

- **パイロット導入**：小規模な実験から開始し、成功事例を蓄積
- **段階的拡大**：臨界点を超えないよう、徐々に適用範囲を拡大
- **デフォルト設定の変更**：現状維持の方向に働く制度上のデフォルトを見直し

6.2.2 確証バイアス：多様な視点の構造的導入

確証バイアスへの対処として、以下の戦略を提案する：

- 「**悪魔の代理人**」の制度化：AIまたは人間による批判的視点の提示を必須化
- **多様なステークホルダーの参画**：異なる立場・利害を持つ主体の関与を確保
- **反証可能性の確保**：仮説を反証する証拠を積極的に探索するプロセス

6.2.3 狹い視野：全体目標の可視化

狭い視野への対処として、以下の戦略を提案する：

- **横断的評価指標**：単一の指標ではなく、複数の指標による評価
- **システムマップの作成**：政策の全体像を可視化した図の共有
- **長期的影響の分析**：短期的な成果だけでなく、長期的な影響も評価

6.3 生成 AI を組み込んだ制度設計

6.3.1 政治的-行政的インターフェース

政治レベルと行政レベルの間で、AI は以下のように位置づけられる：

- **政治的判断**：人間（政治家・議会）が最終的に決定
- **行政的分析**：AI が情報の整理・分析を支援
- **民主的アカウンタビリティ**：人間が説明責任を負う

6.3.2 行政的-事業的インターフェース

行政レベルと事業レベルの間で、AI は以下のように位置づけられる：

- **政策の具体化**：AI が選択肢の生成を支援
- **運用の効率化**：AI が日常的な判断を補助
- **人間による監督**：重要な判断は人間が行う

6.3.3 ZK-SNARKs 型システムの制度的位置づけ

ZK-SNARKs 型政策評価システムは、以下の制度的枠組みの中で運用される：

1. **評価基準の共同設計**：市民討議を通じた基準の策定
2. **透明性の確保**：AI の評価プロセスと基準の公開
3. **控訴の権利**：評価結果に対する異議申立ての機会
4. **人間による最終判断**：AI の評価は参考情報として位置づけ

6.4 「連携・共創」の再設計

6.4.1 三層制度設計（S-T1-T2-O）の再考

STO フレームワークを踏まえ、認知バイアスと生成 AI を考慮した制度設計を提案する：

6.4.2 生成 AI を「杖」として活用するガバナンス

生成 AI を活用するガバナンスのあり方として、以下の原則を提案する：

1. **補完性の原則**：AI は人間の能力を補完し、代替しない
2. **透明性の原則**：AI の活用方法と限界を明示する
3. **アカウンタビリティの原則**：最終的な判断と責任は人間が負う
4. **継続的学習の原則**：AI と人間の協調プロセスを継続的に改善する

Table 6.1: 認知バイアスと生成 AI を考慮した三層制度設計

| 階層 | 主な課題 | 対処戦略 |
|----|-----------------|-------------------|
| S | 現状維持バイアスによる変化抵抗 | 長期ビジョンの共有、段階的目標設定 |
| T1 | 狭い視野による部分最適化 | 横断的指標、全体目標の可視化 |
| T2 | 確証バイアスによる視点の固定化 | 多様な視点の構造的導入 |
| O | 日常的な認知バイアス | AI による対話的介入 |

6.5 小括

本章では、第 4 章の計算論的分析と第 5 章の ZK-SNARKs 型システム設計を踏まえ、制度設計への示唆を導出した。主な貢献は：

1. 認知バイアスへの具体的な対処戦略の提示
2. 生成 AI を組み込んだ制度設計の方向性の提示
3. 「連携・共創」の再設計に向けた枠組みの提示

次章では、本研究の総括と、都市計画への展開について論じる。

Chapter 7

結論：都市計画への展開

7.1 研究の総括

7.1.1 研究目的の達成

本研究は、以下の3つの目的を掲げた：

1. 公共交通政策における実装ギャップの要因として、人間の認知バイアスの影響を計算論的に解明する
2. 生成AIと人間の協調的関係性を具体化するシステムとして、ZK-SNARKs型政策評価システムの可能性を探る
3. 認知バイアスと生成AIを考慮した制度設計への示唆を導出する

これらの目的に対して、本研究は以下の成果を上げた：

目的1について 第4章において、協調ロボット制御モデルを用いた計算論的分析により、三つの認知バイアスが政策協調に与える影響を定量的に解明した。特に、現状維持バイアスの閾値効果、確証バイアスの逆説的効果、狭い視野の一貫した負の影響を発見した。

目的2について 第5章において、ZK-SNARKsの概念を援用した政策評価システムを提案し、LLM as a Judge、Constitutional AI、市民討議を組み合わせたアーキテクチャを設計した。

目的3について 第6章において、認知バイアスへの対処戦略、生成AIを組み込んだ制度設計、「連携・共創」の再設計に向けた枠組みを提示した。

7.1.2 核となる主張

本研究の核となる主張は、以下の通りである：

生成AIは人間の「執政の創造性」を補完する「杖」として、認知バイアスへの対話的介入を通じて政策形成を支援できる。ZK-SNARKsの概念を援用することで、秘匿性と信頼性を両立した政策評価システムが可能になる。

7.2 理論的貢献

7.2.1 認知バイアスの計算論的分析手法の政策科学への導入

本研究は、協調ロボット制御理論を政策ネットワーク分析に応用し、認知バイアスの影響を計算論的に解明する手法を導入した。このアプローチは、政策科学における計算論的転回 (computational turn) の一環として位置づけられる。

7.2.2 ZK-SNARKs 概念の政策評価への応用

本研究は、暗号技術の概念である ZK-SNARKs を政策評価に応用する試みとして先駆的である。この概念転用は、「秘密を守りながら専門性を証明する」という新たな政策参加のあり方を示唆している。

7.2.3 生成 AI の「杖」としての理論的位置づけ

本研究は、生成 AI と人間の関係性を「杖」として補完的に位置づける理論的枠組みを提示した。これは、AI による代替ではなく、AI と人間の協調を前提とする視点である。

7.3 実践的貢献

7.3.1 制度設計への提言

本研究は、認知バイアスと生成 AI を考慮した制度設計への具体的な提言を行った：

- 現状維持バイアスへの対処：段階的变化導入の設計
- 確証バイアスへの対処：「悪魔の代理人」の制度化
- 狹い視野への対処：横断的評価指標の導入
- 生成 AI の活用：補完性、透明性、アカウンタビリティの原則

7.3.2 ZK-SNARKs 型政策評価システムの設計指針

本研究は、ZK-SNARKs 型政策評価システムの具体的な設計指針を提示した：

- 三層アーキテクチャの採用
- Constitutional AI と市民討議による評価基準設計
- 控訴プロセスによる人間介入の確保

7.4 今後の課題：都市計画を舞台にした実証

7.4.1 より複雑な政策領域への適用

公共交通政策は、本研究の「実験場」として適切であったが、より複雑な政策領域への適用が期待される。特に、都市計画は以下の点で興味深い：

- 多様なステークホルダー：土地所有者、開発業者、住民、行政など
- 多様な秘密情報：土地利用計画、開発権、資産価値など
- 長期的影响：数十年単位での都市構造の変化

7.4.2 ZK-SNARKs システムの社会実装

ZK-SNARKs 型政策評価システムの社会実装に向けては、以下の課題がある：

1. 技術的実装：TEE、LLM、Constitutional AI の統合
2. 制度的設計：法的位置づけ、運用ルールの策定
3. 社会的受容：市民の理解と信頼の獲得

7.4.3 生成 AI と人間の協調的関係性の継続的検証

生成 AI 技術は急速に進化しており、人間-AI 協調のあり方も変化し続ける。本研究の枠組みは、継続的な検証と改善を必要とする。

7.5 結び

本研究は、公共交通政策を舞台に、生成 AI と人間の協調的関係性を探求した。その過程で、認知バイアスによる協調失敗のメカニズムを計算論的に解明し、ZK-SNARKs 概念を援用した政策評価システムを提案し、制度設計への示唆を導出した。

公共交通政策は、本研究の「第一の舞台」であった。今後は、都市計画という「第二の舞台」での実証を通じて、生成 AI と人間のより良い協調のあり方を探求していきたい。

参考文献

- [1] Daniel Kahneman. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2011.
- [2] Ben Shneiderman. *Human-Centered AI*. Oxford University Press, 2020.
- [3] Hidehiko Yoshihara et al. “Cooperative Control of Multi-Joint Robot Arm”. In: *Proceedings of IEEE International Conference on Robotics and Automation*. 2009, pp. 1–6.
- [4] Chris Ansell and Alison Gash. “Collaborative Governance in Theory and Practice”. In: *Journal of Public Administration Research and Theory* 18.4 (2008), pp. 543–571.
- [5] Hirokazu Kato et al. “Community-Participatory Regional Public Transport”. In: *Transport Policy Studies* 12.2 (2009), pp. 2–9.
- [6] Kirk Emerson, Tina Nabatchi, and Stephen Balogh. “An Integrative Framework for Collaborative Governance”. In: *Journal of Public Administration Research and Theory* 22.1 (2012), pp. 1–29.
- [7] William Samuelson and Richard Zeckhauser. *Status Quo Bias in Decision Making*. Vol. 1. 1. 1988, pp. 7–59.
- [8] Federica Russio et al. “Confirmation Bias in Policy Decision Making”. In: *Policy Sciences* 48.3 (2015), pp. 289–313.
- [9] Eli Ben-Sasson et al. “Succinct Non-Interactive Zero Knowledge for a von Neumann Architecture”. In: *Proceedings of USENIX Security Symposium*. 2014, pp. 781–796.
- [10] Ukyo Nagata. “Is ”Japan MaaS” a Type of ”Transport Urban Planning”? A Study on Goals and Governance”. In: *Japan Evaluation Research* XX.X (2024), pp. XX–XX.
- [11] Didier Van de Velde. *Organisational Forms and Entrepreneurship in Public Transport*. Thesis Publishers, 1999.
- [12] Santo Fortunato. “Damage Spreading and Opinion Dynamics on Scale Free Networks”. In: *Physica A* 348 (2005), pp. 683–690.

Appendix A

付録

A.1 協調ロボット制御モデルの数式展開

A.1.1 運動学モデル

N関節ロボットアームにおいて、各関節の位置は以下の式で与えられる：

$$\mathbf{x}_i = \mathbf{x}_{i-1} + a_i \begin{bmatrix} \cos(\sum_{j=0}^{i-1} \theta_j) \\ \sin(\sum_{j=0}^{i-1} \theta_j) \end{bmatrix} \quad (\text{A.1})$$

ここで、 $\mathbf{x}_0 = \mathbf{0}$ である。

A.1.2 協調係数の計算

各関節の協調係数 k_i は、以下の式で計算される：

$$k_i = \exp \left(-4 \ln(2) \frac{\|\mathbf{v}_{l,i} - \mathbf{v}_{i+1}\|^2 + \epsilon_1}{\|\mathbf{v}_{l,i}\|^2 + \epsilon_2} \right) \quad (\text{A.2})$$

ここで、 ϵ_1, ϵ_2 は数値安定性のための小さな正の定数である。

A.2 Japan MaaS プロジェクト分析の詳細

A.2.1 分析対象プロジェクト一覧

2020 年度に認定された 38 の Japan MaaS プロジェクトを分析対象とした。プロジェクトは以下のカテゴリに分類される：

- 観光型 MaaS : XX 件
- 地域課題解決型 MaaS : XX 件
- 企業主導型 MaaS : XX 件

A.2.2 評価指標のコーディング基準

評価指標は以下の基準でコーディングした：

Table A.1: 評価指標のコーディング基準

| カテゴリ | コーディング基準 |
|-------|------------------------|
| 事業指標 | 利用者数、収益、運行効率など |
| 非事業指標 | 社会影響、環境効果、アクセシビリティ改善など |
| 市民参加 | ワークショップ、アンケート、協議会など |

A.3 ZK-SNARKs型政策評価システムの実装詳細

A.3.1 システム構成

ZK-SNARKs型政策評価システムは、以下のコンポーネントから構成される：

- 秘匿化処理モジュール
- LLM 評価エンジン
- Constitutional AI 基準ライブラリ
- 控訴処理インターフェース

A.3.2 技術的仕様

Table A.2: 技術的仕様

| 項目 | 仕様 |
|------------------|--------------------|
| LLM | GPT-4 / Claude |
| Temperature | 0 |
| Self-Consistency | 5 回 |
| TEE | AWS Nitro Enclaves |