

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	xi
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 「執政の創造性」とは何か	3
2.2.1 定義	3
2.2.2 政策過程モデルにおける価値判断	3
2.2.3 ウィキッド・プロブレムとの関連	4
2.3 ルーマン理論による基礎づけ	4
2.3.1 なぜルーマンか	4
2.3.2 コミュニケーションの3段階：情報・伝達・理解	5
2.3.3 構造的カップリング：本研究の核心概念	5
2.3.4 「理解の創造性」の定義	5
2.3.5 AI の原理的限界	6
2.4 Human-AI Policy：政策形成における AI と人間の関係性	6
2.4.1 AI 政策論の展開	6
2.4.2 Human-Centered AI の理念	6
2.4.3 AI の「杖」としての位置づけ	7
2.5 協調制御理論と社会システムへの応用	7
2.5.1 協調制御理論の基礎	7
2.5.2 社会システムへの応用可能性	8
2.6 協調的ガバナンス論	8
2.6.1 協調的ガバナンスの定義	8
2.6.2 公共交通における連携・共創	8
2.6.3 実装ギャップの指摘	8

2.7	認知バイアスと意思決定	8
2.7.1	行動経済学の基礎概念	8
2.7.2	政策プロセスにおける認知バイアス	8
2.8	ZK-SNARKs と政策評価	9
2.8.1	ZK-SNARKs の基本概念	9
2.8.2	政策評価への応用可能性	9
2.9	小括：理論的空白の特定	9
3	舞台としての公共交通政策：現状と課題	10
3.1	はじめに	10
3.1.1	本章の目的	10
3.1.2	MaaS の勃興	10
3.1.3	先行研究	10
3.1.4	リサーチクエスチョンと仮説	11
3.2	MaaS の概念と展開	11
3.2.1	フィンランドにおける MaaS の勃興	11
3.2.2	MaaS に必要な統治構造変革	12
3.2.3	MaaS とインターネットの類似性	12
3.2.4	日本版 MaaS の定義と展開	12
3.3	法体系と交通まちづくり	13
3.3.1	交通政策基本法	13
3.3.2	交通政策基本計画	13
3.3.3	地域公共交通活性化・再生に関する法律	13
3.3.4	交通まちづくりの定義とその親和性	13
3.4	市民参加の捉え方	14
3.4.1	公共交通政策における市民参加に関する議論	14
3.4.2	公共サービスとしての公共交通	14
3.4.3	望ましい市民参加のスタンス	14
3.5	実装ギャップの実証分析：Japan MaaS	14
3.5.1	分析手法	14
3.5.2	分析対象	15
3.5.3	分析結果	15
3.5.4	事例分析	15
3.5.5	考察	16
3.5.6	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	16
3.6	小括：公共交通政策を「実験場」として位置づける理由	17
4	認知バイアスの政策協調への影響：計算論的分析	18
4.1	研究背景と問題設定	18
4.1.1	研究の背景	18
4.1.2	政策の文脈と変遷	18
4.1.3	実装ギャップの実証的証拠	18
4.1.4	研究目的とアプローチ	19
4.2	理論的枠組み	19
4.2.1	協調ガバナンス理論	19
4.2.2	公共サービス提供における共創	20
4.2.3	協調的政策実装における認知バイアス	20
4.2.4	協調制御理論とマルチエージェント調整	20
4.3	計算論的モデリング・フレームワーク	20
4.3.1	類推的マッピング：正当化と限界	20
4.3.2	モデル構造	21
4.3.3	運動学モデル	21

4.3.4 制御目標と調整メカニズム	22
4.4 認知バイアスの統合	22
4.4.1 現状維持バイアスのモデル化	22
4.4.2 確証バイアスのモデル化	22
4.4.3 狹い視野のモデル化	22
4.5 シミュレーション実験	23
4.5.1 実験設計	23
4.5.2 評価指標	23
4.6 結果	23
4.6.1 確証バイアスの効果	23
4.6.2 現状維持バイアスの閾値効果	23
4.6.3 狹い視野の一貫した負の影響	24
4.7 統計的サマリー	24
4.8 制度的設計への示唆	24
4.9 小括：認知バイアスによる協調失敗のメカニズム	24
5 生成 AI と人間の関係性：ZK-SNARKs 型政策評価システム	25
5.1 はじめに：ウィキッド・プロブレムと政策評価の課題	25
5.1.1 ウィキッド・プロブレムの台頭	25
5.1.2 コミュニケーションにおける二重の困難	25
5.2 ZK-SNARKs 概念の援用	25
5.2.1 ZK-SNARKs とは何か	25
5.2.2 ZK-SNARKs が実用的な理由	26
5.2.3 秘密を守りながら専門性を証明する	26
5.3 ZK-SNARKs 型政策評価フレームワークの提案	26
5.3.1 フレームワークの目的	26
5.3.2 入札型秘匿証明の構築	27
5.4 LLM as a Judge による実装	27
5.4.1 LLM as a Judge と ZK-SNARKs の親和性	27
5.4.2 LLM as a Judge の可能性	27
5.4.3 LLM as a Judge の限界と課題	27
5.5 LLM as a Judge の限界を超えるための人間支援機能	28
5.5.1 Constitutional AI と市民討議	28
5.5.2 計画・推論分離型システム	28
5.5.3 控訴プロセスによる人間介入	28
5.5.4 決定論的運用	29
5.6 ZK-SNARKs 型政策評価のアーキテクチャ	29
5.6.1 三層アーキテクチャ	29
5.6.2 生成 AI の「杖」としての位置づけ	29
5.6.3 限界と課題	30
5.7 小括	30
6 制度設計への示唆	31
6.1 理論的含意	31
6.1.1 第 4 章のシミュレーション結果からの設計原則	31
6.1.2 第 5 章の ZK-SNARKs システムからの制度的含意	31
6.2 各バイアスへの対処戦略	31
6.2.1 現状維持バイアス：小さな変化の積み重ね	31
6.2.2 確証バイアス：多様な視点の構造的導入	31
6.2.3 狹い視野：全体目標の可視化	32
6.3 生成 AI を組み込んだ制度設計	32
6.3.1 政治的-行政的インターフェース	32

6.3.2 行政的-事業的インターフェース	32
6.3.3 ZK-SNARKs型システムの制度的位置づけ	32
6.4 「連携・共創」の再設計	32
6.4.1 三層制度設計（S-T1-T2-O）の再考	32
6.4.2 生成AIを「杖」として活用するガバナンス	32
6.5 小括	33
7 結論：都市計画への展開	34
7.1 研究の総括	34
7.1.1 研究目的の達成	34
7.1.2 核となる主張	34
7.2 理論的貢献	34
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	34
7.2.2 ZK-SNARKs概念の政策評価への応用	35
7.2.3 生成AIの「杖」としての理論的位置づけ	35
7.3 実践的貢献	35
7.3.1 制度設計への提言	35
7.3.2 ZK-SNARKs型政策評価システムの設計指針	35
7.4 今後の課題：都市計画を舞台にした実証	35
7.4.1 より複雑な政策領域への適用	35
7.4.2 ZK-SNARKsシステムの社会実装	35
7.4.3 生成AIと人間の協調的関係性の継続的検証	36
7.5 結び	36
A 付録	38
A.1 協調ロボット制御モデルの数式展開	38
A.1.1 運動学モデル	38
A.1.2 協調係数の計算	38
A.2 Japan Maasプロジェクト分析の詳細	38
A.2.1 分析対象プロジェクト一覧	38
A.2.2 評価指標のコーディング基準	38
A.3 ZK-SNARKs型政策評価システムの実装詳細	39
A.3.1 システム構成	39
A.3.2 技術的仕様	39

List of Figures

List of Tables

3.1	Japan MaaS 36 プロジェクトの評価指標分析	15
4.1	Japan MaaS 38 プロジェクトの実装ギャップ分析	19
4.2	シンボル-概念対応表	21
4.3	現状維持バイアスの閾値効果	23
4.4	認知バイアス効果の統計的サマリー	24
6.1	認知バイアスと生成 AI を考慮した三層制度設計	33
A.1	評価指標のコーディング基準	39
A.2	技術的仕様	39

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、変化への抵抗や現状の維持選好を示す現状維持バイアス (Status Quo Bias)、自己の信念を確認する情報の優先的選択という確証バイアス (Confirmation Bias)、局所的最適化への固執や全体最適の見落としという狭い視野 (Narrow Framing) が重要な影響を及ぼ

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

す。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。第一に、規範的判断・価値創造の不在である。AI は何が社会にとって「善い」のかを判断できない。第二に、文脈理解の困難性である。学習データの範囲外の「未知の状況」への適応には限界がある。第三に、「創造性」の源泉の欠如である。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。第一に、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。第二に、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。第三に、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 7 章から構成される。第 2 章では先行研究のレビューを行い、Human-AI Policy の議論を中心に整理する。第 3 章では舞台としての公共交通政策の現状と課題を論じる。第 4 章では認知バイアスの政策協調への影響について計算論的分析を行う。第 5 章では生成 AI と人間の関係性として、ZK-SNARKs 型政策評価システムを提案する。第 6 章では制度設計への示唆を導出する。第 7 章では結論として、都市計画への展開について論じる。

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、「執政の創造性」とルーマン理論、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 と政策評価の六つの領域を中心に整理し、理論的空白を特定する。

2.2 「執政の創造性」とは何か

2.2.1 定義

「執政の創造性」とは、以下のように定義される：

執政の創造性：社会の構成員同士のコミュニケーションを前提に、価値判断を基盤として、新たなガバナンスの範囲を生成・配分していく能力。

この定義は、三つの要素から構成される。第一に、政策は、単一の主体による合理的決定ではなく、複数の主体間のコミュニケーションを通じて形成される。第二に、政策決定は統計的平均や客観的指標への還元が不可能な、主体的な価値判断を要する。第三に、政策は固定された枠組みではなく、状況に応じて新たな範囲を創造し続ける動的なプロセスである。

この概念は、カンギレム (Canguilhem) の「規範の創造」概念 [2] および井庭の創造システム理論 [3] を統合したものである。カンギレムは、「正常」とは統計的平均ではなく、「生の独自的な規範性」を肯定することであると論じた。健康とは客観的指標ではなく、主体によって体験される価値である。同様に、政策における「正しさ」も、統計的効率性ではなく、主体による価値判断に基づく。

井庭の創造システム理論では、コミュニケーションの「わかり合えなさ」から出発し、「発見」を要素とするシステム上で創造が起こるとされる。本研究は、この創造のプロセスを政策の文脈に適用し、「執政の創造性」として定式化する。

2.2.2 政策過程モデルにおける価値判断

「執政の創造性」の重要性は、既存の政策過程モデルを分析することで裏付けられる。主要な政策過程モデルを見てみよう。

政策の窓モデル

キングドン [4][5] による政策の窓モデルは、政策過程に「問題」「政策代替案」「政治」という 3 つの独立した流れが存在し、それらが特定の時点で合流したときに政策決定が生じることを示す。

キングドンは、この3つの流れが合流する瞬間を「政策の窓」が開くと表現する。窓は一時的にしか開かず、逃したら再び開くまで長く待たなければならない。

このモデルが示すのは、政策決定は論理的な手順ではなく、問題と解決案と政治的機会の偶然の出会いによって生じるということである。この「偶然の出会い」を見極め、窓が開いた瞬間に行動するには、人間の価値判断と政治的勘が不可欠である。AIは過去のデータから傾向を分析することはできても、「今がチャンスだ」と感じ取り、その瞬間に動くという状況認識と行動のタイミングについては、人間の判断に依存せざるを得ない。

唱導連携モデル

唱導連携モデル（Advocacy Coalition Framework）は、サバティエ^[6]らによって開発されたモデルであり、特定の政策分野において、共通の信念体系を共有する主体たちが連携（コアリション）を形成し、政策を唱導するプロセスを分析する。

このモデルの特徴は、政策を単なる利害調整ではなく、信念体系の競合として捉える点にある。各連携は、(1) 深層核信念（基本的価値観）、(2) 政策核信念（具体的政策目標）、(3) 二次的側面（手段的判断）という階層的な信念体系を持つ。

ここで重要なのは、どの信念を優先すべきか、どの連携の主張を採用すべきかという判断が、常に価値判断を伴うということである。統計的データや客観的分析だけでは、信念の競合を解決できない。どの価値を重視するかという判断が必要であり、これは人間の創造的適応能力に依存する。

村松モデル

村松^[7]モデルは、日本の政策過程を「与党・官僚・利益団体」の三者関係として分析する枠組みである。このモデルでは、政策決定はこれら三者の交渉と取引を通じて行われる。

真渕による修正版では、野党勢力や新規参入者の役割も考慮される。既存の産業組織が維持しようとする価値と、新規参入者が求める変革の価値が衝突する中で、政治家はバランスを取る必要がある。

このモデルが示すのもまた、利害の調整と価値の優先順位付けが政治的本質であり、それは計算可能な最適化問題ではないということである。三者間の「妥当な落としどころ」を見出すには、人間同士の交渉と相互調整が不可欠である。

2.2.3 ウィキッド・プロブレムとの関連

公共政策の多くは、Rittel & Webber^{[8][9]}の「ウィキッド・プロブレム（Wicked Problems）」の性質を帶びている。ウィキッド・プロブレムとは、決定的な問題定義がなく、問題そのものが何であるかについて利害関係者の間で合意がない問題である。停止ルールもなく、いつ問題が「解決」されたのかを客観的に判断できない。解は「良い/悪い」ではなく「より良い/より悪い」であり、最適解は存在せず、複数の利害のバランスを取るしかない。また、解を適用した結果、予期せぬ副作用が生じる可能性があり、社会実験はやり直しが効かない。さらに、すべてのウィキッド・プロブレムは本質的にユニークであり、過去の経験から単純に適用できない。

ウィキッド・プロブレムへの対応は、終わりのない創造的プロセスである。そして、「終わりがないからこそ、政治がその不快さを受け入れる必要がある」のである。この終わりのない創造的プロセスこそが、人間による政策の固有の意味であり、AIには代替不可能な領域である。

2.3 ルーマン理論による基礎づけ

2.3.1 なぜルーマンか

コミュニケーションを理論的基軸とする場合、ユルゲン・ハーバーマスの「コミュニケーション的行為理論」^[10]も選択肢として存在する。しかし、本研究はルーマン^[11]の立場を採用する。その理由を説明する。

ハーバーマスは、コミュニケーションを「了解志向的」な行為として捉える。コミュニケーションの理想的状態では、参加者が互いに「了解」に到達し、合意を形成することが期待される。こ

の立場からは、コミュニケーションの「失敗」や「誤解」は、理想的言語状況からの逸脱として問題視される。

これに対し、ルーマンはコミュニケーションを「了解の否定」を内包するプロセスとして捉える。ルーマンによれば、コミュニケーションは常に「理解」と「誤解」の双方を可能性として含んでおり、「わかり合えなさ」こそがコミュニケーションの本質的な特徴である。

本研究がルーマンの立場を採用する理由は、以下の二点にある。

第一に、「わかり合えなさ」を「創造の源泉」として位置づける視点が必要である。ハーバーマス的な「理想的了解」を前提とすれば、AIも「十分に良い」情報処理を行うことで「機能的な了解」に貢献できると主張しうる。しかし、ルーマン的な視点からは、「わかり合えなさ」から生じる価値判断と意味構成のプロセスこそが創造の核心であり、これをAIは代替できない。

第二に、政策的決定における「価値競合」の不可避性も重要である。公共政策は、ハーバーマスが想定する「理想的言語状況」において理性の力だけで解決可能な問題ではなく、複数の正当な価値が競合するウィキッド・プロブレムとしての性質を持つ。このような状況では、「了解」への到達よりも、「わかり合えなさ」を前提とした創造的適応こそが求められる。

2.3.2 コミュニケーションの3段階：情報・伝達・理解

ルーマン[11][12]によれば、コミュニケーションは「情報(Information)・伝達(Mitteilung)・理解(Verstehen)」という3つの選択過程から構成される。

情報(Information) 多数の可能性の地平からの一つの選択であり、何を語るかの選択である。

伝達(Mitteilung) 多数の伝達可能性からの選択であり、いかに語るかの選択である。

理解(Verstehen) 多数の理解可能性からの選択であり、いかに受け止めるかの選択である。

ルーマンは、「三つの選択のはたらきのすべてが総合されるときにはじめてコミュニケーションというものが成り立つ」と強調する。この3層構造は、コミュニケーションが単純な情報伝達ではなく、各段階で選択と解釈が介在する複雑な過程であることを示している。

2.3.3 構造的カップリング：本研究の核心概念

ここで重要なのが、ルーマンにおける「構造的カップリング(strukturelle Kopplung)」[11][13]の概念である。これは、作動上は完全に独立(閉鎖)している複数のシステムが、互いに不可欠な環境として影響し合う関係を指す。

ルーマン理論において、最も重要な構造的カップリングは、「心的システム(意識)」と「社会システム(コミュニケーション)」の関係である。心的システムは思考し、社会システムはコミュニケーションする——それぞれ別の作動を行う。しかし、心的システムがなければコミュニケーションは発生しない。意識はコミュニケーションに対して、刺激や誘発を与えたり、あるいは邪魔をしたりすることができる。ただし、意識がコミュニケーションを「因果的に決定」するわけではない。意識はあくまで環境として、コミュニケーション・システムに「刺激」を与え、システム側がそれを独自の論理で処理する¹。

2.3.4 「理解の創造性」の定義

以上の理論的整理を踏まえ、本研究で論じる「理解の創造性」を以下のように定義する：

理解の創造性：心的システムと社会システムの構造的カップリングにおいて、心的システムから社会システムへの刺激として提供される、価値判断を伴う意味構成のプロセス。

¹ここで注意すべきは、ルーマン理論において「認知」は心的システム内部に所在し、社会的システムに「分布」しているわけではないことである。構造的カップリングは、社会的コミュニケーションが心的システムを刺激することを可能にするが、認知プロセスそのものは心的システムのオートポイエーシスとして完結する。

この定義は、ルーマンの厳密な意味での「理解」（コミュニケーション接続）とは区別される。本研究が着目するのは、コミュニケーション接続を駆動する「価値判断の源泉」としての心的システムの役割である。

この観点から、「わかり合えなさ」は単なる誤解ではなく、構造的カップリングにおいて各心的システムが独自の価値判断に基づいて意味を構成する結果として生じる、コミュニケーションの本質的な特徴として理解される。

この価値判断を伴う意味構成のプロセスこそが、人間固有の創造性であり、本研究が「執政の創造性」として定式化する対象である。

2.3.5 AI の原理的限界

ルーマンの理論において、機械はオートポイエシス・システム（生命・意識・社会）とは区別される「非ポイエティック」な存在とされている [11]。AI システムも基本的には計算機プログラム（機械）上で動作するシステムである。AI——その最先端の形態を含めて——は心的システムに該当しない。

オートポイエティックでない：たとえ自己学習能力を持つ AI であっても、その「学習」は人間が設計したアルゴリズムと訓練データに依存している。外部からの入力なしに自律的に自身の「思考」を産出し続ける閉鎖的なネットワークを持たない。

自己言及的な意味処理を行わない：AI の確率的出力は、文脈に応じて「もっともらしい」次のトークンを選択するが、この選択プロセスは自己言及的ではない。AI は「この意味を選択したこと自体」を次の処理の地平として開くことはなく、単に統計的パターンに基づいて出力を生成する。

価値判断を伴わない：心的システムからの刺激は「何が重要か」「何を優先すべきか」という規範的判断を前提とするが、AI の出力は統計的パターンに基づいており、独自の規範的判断を伴わない。以上の議論から、公共政策における AI の原理的限界が明らかになった。AI は「情報」の選択や「伝達」の補助には機能しうるが、「理解」の創造的プロセスには原理的に参加できない。AI は心的システムを持たないため、価値判断を伴う意味構成を行えないのである。

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

2.4.1 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 [14]。

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。

2.4.2 Human-Centered AI の理念

Shneiderman (2022) [14] 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。

2.4.3 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。

この視点からは、いくつかの設計原則が導かれる。第一に、AIは人間の最終判断を前提とする事である。第二に、AIの限界を明示的に理解することである。第三に、人間-AI協調のプロセスを透明化することである。第四に、AI自身のバイアスに対処することである。第五に、説明責任は常に人間が負うことである。

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。

生成AIは、いくつか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、そして認知バイアスの指摘（「悪魔の代理人」機能）などがその例である。

一方で、いくつか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、そして最終的な責任の所在などがその例である。

2.5 協調制御理論と社会システムへの応用

2.5.1 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 [15]。

この理論は、いくつかの重要な特徴を持つ。分散的な意思決定を行うこと、局所的な情報に基づく協調を行うこと、そして全体的な目標の達成を目指すことである。

2.5.2 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）。

2.6 協調的ガバナンス論

2.6.1 協調的ガバナンスの定義

Ansell and Gash (2008) [16] 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」

2.6.2 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 [17]。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、関係ステークホルダー間での認識と責任分担の共有、各ステークホルダーが参加から利益を得られること、ステークホルダーを調整するキーパーソンの存在、そして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を指摘している。

2.6.3 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) [18] は、協調的ガバナンスが直面する課題として、高い取引コスト、最小公約数的な解決策への収束、そして組織された利益による捕捉を指摘している。

2.7 認知バイアスと意思決定

2.7.1 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。

2.7.2 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) [19] 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 [20]。

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]。

2.8 ZK-SNARKs と政策評価

2.8.1 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 [21]。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が一切漏洩しない

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

2.8.2 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）。

2.9 小括：理論的空白の特定

先行研究のレビューから、いくつかの理論的空白が明らかになった。第一に、「執政の創造性」の理論化である。ルーマンの社会システム理論を用いて「執政の創造性」を基礎づけ、AIと人間の役割分担を明確にした研究は限定的である。第二に、認知バイアスと政策協調の接続である。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。第三に、計算論的分析手法の欠如である。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。第四に、ZK-SNARKs 概念の政策評価への応用である。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

3.1.1 本章の目的

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举し、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。

3.1.2 MaaS の勃興

近年、日本において「Mobility as a Service」の勃興が著しい。2019 年度より始まった「日本版 MaaS 推進事業」は、瞬く間に全国各地に広がり、現在までに 100 事業以上のプロジェクトを送り出している。本稿ではこうしたプロジェクトがどのような体制の下に進められ、またどのような検証体制が敷かれているかについて、国土交通省の公開している資料を基として実証的に考察する。地域公共交通政策のうち、Mobility as a Service（MaaS）と呼ばれる分野は、この数年で急速に勃興してきた。それ以前から、「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」として、「交通まちづくり」という思想体系が存在、発展してきた。

3.1.3 先行研究

日本版 MaaS について、その特徴を述べた文献はいくつかある。

まず注目されるのは、特に過疎地域や地方中核市における交通が抱える課題を解決するためのアプローチとしての見立てである。魏（2022）は、日本の MaaS の注目点が「地方型 MaaS」である点に言及したうえで、日本の MaaS の特徴について以下のように述べている。

「地方型」MaaS は、地域の資源や特性を最大限活かしながらより満足度の高いモビリティサービスを提供することであり、地域にある交通モード間の連携強化により、その場その場で発生した交通需要（生活交通需要、観光交通需要）を上手く汲み上げ、集約することで実現できるものである。

また藤本（2021）では、欧州における MaaS が「都市型 MaaS」と「地方型 MaaS」に分化されており、また日本版 MaaS が「地方型」を含む 5 類型に分割されている点に触れたうえで、地方型 MaaS を検討するポイントを 5 つに整理している。

MaaS 自体の実装にかかる研究も進んでいる。後藤（2021）は、MaaS を旅行業の一類型ととらえ、企画旅行業務を適用する可能性やデータ供給の課題について論じている。

既存の公共交通運営体制に関する議論は、多くの場合「どのような取り組み」があったか、また「どのような評価が可能か」という 2 つに大別される。前者の場合、生活バスよっかいちについての事例報告を行った福本・加藤（2012）や、多くの日本モビリティマネジメント会議の報告に代表される。また「どのような評価が可能か」については、アクセシビリティ評価の地域への適用について実例を分析した喜多（2022）などがあげられる。公共交通に関する責任分担や意思決定組織についての検討も複数ある。

本稿で前提とした「交通まちづくり」の考えでは、交通政策における市民参加と地域戦略との連携が重視されていると考えられる。また後述するものであるが、地域公共交通に関する法体系においても、交通政策の戦略性を重視している。ただ少なくとも「日本版 MaaS」において、法体系からの課題意識を導入した包括的な研究、また交通事業者としてではない形で計画の要件定義から携わる市民参画の有無、に関する実証的な研究は見当たらなかった。

3.1.4 リサーチクエスチョンと仮説

本章は、いわゆる「日本版 MaaS」を「交通まちづくり」実践の一類型と捉え、以下の 2 つの問い合わせについて検討し、結果としてこの捉え方が適切かについて確認するものである。第一に、「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。第二に、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である。

本分析における仮説としてはそれぞれ、目標設定は MaaS 事業の事業性に関して設定されるがそれ以外の地域に関する効果については設定されないこと、そして市民による批判回路は協議会に閉じずパブリックコメント・市民代表参加・座談会・ワークショップなどの機会を通じて設定されていること、と設定している。

3.2 MaaS の概念と展開

3.2.1 フィンランドにおける MaaS の勃興

まず MaaS は、Hekkilä（2014）にて初めて学術的な表舞台に立った、公共交通の新しい提供概念である。氏が所属する大学へ依頼されたヘルシンキ市の交通課題解決のために、修士論文で書き上げたものである。これによると、Nokia を主軸としたフィンランドの通信政策やその事業に対する理解の蓄積が、ほぼそのまま交通政策に導入されていることがわかる。

この論文では、論文冒頭のサマリーにおいてその目的を明確に示している。具体的には、公共交通の利用促進が叫ばれ、都市構造への影響が避けられない中で、既存の交通サービス目標は「課題に十分にこたえられていない」と批判している。そのうえで、「公共交通セクターの構造転換が必要である」として「多様で魅力的な交通サービスの便利な提供」として MaaS を定義している。具体的な MaaS の構造として、「交通サービスプロバイダーから交通サービスを買い、それをユーザーへ提供する」主体を「モビリティオペレーター」と定義している。結果として、交通サービスを従来提供していた事業者と利用者の間に、もう一つの主体が挟まることが特徴である。

3.2.2 MaaS に必要な統治構造変革

この MaaS コンセプトを実現するうえで、政策上で何が必要であるのかについて、Heikkilä は 7 つの政策目標を提示している。企業・当局・機関・ユーザーなどすべてのステークホルダーの協力体制の調整、サービス・エコシステムの前提条件を満たすための法律・規制の改正、共通のルールと適切な規制の構築およびその遵守状況の監視、モビリティサービス提供の再編成、変革されたオペレーションの確立、購入・補助金手続きの見直し、そしてパイロットとテストエリアの設置である。

このような政策目標を実現するために、フィンランドでは多様な法改正が行われている。「モビリティ・ローミング」の実現を目指として、既存サービス事業者及び参入する交通事業者が交通に関する情報提供を義務付けられるとともに、チケット発行と決済に関するシステムへの相互運用が可能となる法律が成立、2017年に施行された。

3.2.3 MaaS とインターネットの類似性

そもそもフィンランドでは、通信産業と運輸産業に関する規制を、同一の機関が担当している。インターネットの特徴は、「いかなる情報であっても運搬可能とすること」、そのために「デジタル化」を挟むことである。

フィンランドにて定義された MaaS に存在するインターネットエコシステムと類似の概念は 3 点ある。第一にモジュール化であり、様々な機能についてルールを決め目的ごとに分化して整備させることで、システム上でのメタボリズムと堅牢性を確保する。第二にネットワークであり、前述のモジュールを接続することで一つの目的を完結させることを支援する。第三にインターフェースの統合であり、様々なサービスがあったとしてもユーザーとの接点については統合して管理できるようにする。

3.2.4 日本版 MaaS の定義と展開

日本における、中央官庁による MaaS 推進事業は、主に国土交通省と経済産業省が所管して展開されてきた。2019 年には経済産業省と共同で「スマートモビリティチャレンジ」が開始され、全国 28 事業が指定されることになった。さらに国土交通省単独でも、2019 年より「日本版 MaaS 推進事業」が推進され、初年度においては 19 事業、2020 年度には 36 事業が選定された。

「日本版 MaaS」の定義については、国土交通省により以下のように述べられている：

日本版 MaaS は、地域住民や旅行者一人一人のトリップ単位での移動ニーズに対応して、複数の公共交通やそれ以外の移動サービスを最適に組み合わせて検索・予約・決済等を一括で行うサービスであり、観光や医療等の目的地における交通以外のサービス等との連携により、移動の利便性向上や地域の課題解決にも資する重要な手段となるものです。

この定義を参考すると、日本における MaaS は構造的なものではなく、地域課題解決ベースのサービス提供について重きを置いている。これに基づけば日本版 MaaS は、交通サービスの対価支払い・生産・提供の一連の消費活動を組み替える意図は持たない。つまり、国土交通省が考える MaaS は、制度やシステムの変革というよりもむしろ、交通事業者が提供する公共交通サービスについての改善アイデアコンテスト並びに実証実験、ととらえた方が説明はつきやすい。

3.3 法体系と交通まちづくり

3.3.1 交通政策基本法

交通政策基本法は、2013 年に可決・成立した、日本における交通政策の理念を定めた法律である。本法では、2 条から 7 条までを理念として、国には「国民の理解を得るよう努める」とこと、地方自治体には「その地方公共団体の区域の自然的経済的社会的諸条件に応じた施策を策定し、及び実施する責務」、交通事業者は「その業務を適切に行うよう努めるとともに、国又は地方公共団体が実施する交通に関する施策に協力するよう努める」ことをそれぞれ求め、またこの 3 者ならびに住民その他の関係者に対して、「基本理念の実現に向けて、相互に連携を図りながら協力するよう努める」としている。

公共交通に関わる点の理念について確認すると、「交通の機能を管理する」という点に重点が置かれている。まず第 2 条にて「(交通の) 機能が十分に発揮されることにより、国民その他の者の交通に対する基本的な需要が適切に充足されることが重要であるという基本的認識の下に行われなければならない」として、交通政策の意義が定義されている。

そして第3条にて、目的志向型の交通政策目標設定義務、つまり「交通が、豊かな国民生活の実現に寄与するとともに、我が国の産業、観光等の国際競争力の強化並びに地域経済の活性化、地域社会の維持及び発展その他地域の活力の向上に寄与するものとなるよう、その機能の確保及び向上が図られることを旨として行われなければならない」が記載されている。

3.3.2 交通政策基本計画

交通政策基本計画は、交通政策基本法に基づいて策定される計画である。第1次の計画は2015年に策定され、2021年に更新されている。

計画の構成としては、「交通に関する施策の『基本的方針』、計画期間内に目指すべき『目標』、目標の各々について取り組むべき『施策』の三層構造となっており、関係者の責務・役割や連携・協働等についても、施策の推進に当たって『留意すべき事項』として整理している」。

横ぐしの法律を背景としているが、地域公共交通に関連して目標となっている数値はほとんどの場合、事業がどの程度導入されているかが指標となっている。

3.3.3 地域公共交通活性化・再生に関する法律

地域公共交通活性化・再生に関する法律は、2007年より施行された公共交通における「総合計画法」である。この法律は、元々は「施策支援法」であった。「頑張る地域を応援する」、「LRT」をはじめとした道路運送法だけでは対応できなかったバス事業以外の交通事業について包括的な連携を図るとともに、交通システムごとの支援メニューを取りそろえたシステムであった。

しかし、2010年近傍のいわゆる「事業仕分け」を経て、地域の生活交通を確保する「総合計画法」に転化し、取り組み支援型の事業支援システムを残したまま強制力を持ち始めた。この法律に基づいて策定される「地域公共交通計画」「地域公共交通利便増進計画」を所管する地方運輸局には、路線届出の最終的な受理権限もある。

3.3.4 交通まちづくりの定義とその親和性

主に交通政策基本法を基底とした理念体系を確認すると、この内容が驚くほど「交通まちづくり」のそれと類似していることが分かる。

交通まちづくりは、都市の課題解決のために、公共交通を中心とした交通を動員する交通政策の概念である。既存の交通需要を満たすための交通政策とは異なり、地域の課題解決に貢献する交通を、目的の設定から市民参加を加える形で設計、提供する計画手法で、2005年ごろから主に札幌、横浜、岡山、熊本など政令市における市民団体より勃興してきた。

3.4 市民参加の捉え方

3.4.1 公共交通政策における市民参加に関する議論

公共交通政策における市民参加に関する規範的な議論は、帰結主義に立って市民参加が情報提供と労力提供の役割を持つと説く太田（2009）、非帰結主義に立って6つの要件を示す太田（2000）、実証的に7つの要件を提示し、協働を通じて「新たな公共、地球の公益」を創造すべきとした森栗（2009）などがある。

近年の交通政策における市民参加に関する実証的な議論では、その市民参加形態は、主に不足する財源や人的資源を確保するため、市民が交通運営者として参画する物に集中している。ただ、こうしたスタンスではあくまで、どのように交通事業を成立させるか、また「目の前の交通に対する課題をどのように解決するか」という観点に基づいている。

こうしたスタンスは、こと「横ぐし」の領域を扱う地域政策においては、課題の相対化を妨げ、重要度の高い解決すべき課題を見落とす懸念がある。さらに交通事業者などの公益的な企業は、地域政策の構成員として「選ばれる」というプロセスを経ないことが多いため、概して一方的になりやすい可能性がある。

3.4.2 公共サービスとしての公共交通

一方で、地域公共交通は自治体の計画によって定義され、非競合性を持たないため、インフラとしてではなく公共サービスとして定義されるべきである。通信サービスや道路などの非競合性をある程度持つ財とは違い、利用できる時間や容量が、物理的に決定されているからである。つまり、一旦整備したところで供給量は簡単に変化させられない即時材であるとともに、また提供に必要な労力や資源には限りがある。市民参加を通じてそうした資源を導入できるとしても限界があるから、得られる効果とコストとのバランスについて検討すべきである。

3.4.3 望ましい市民参加のスタンス

以上を踏まえ、市民参加の類型としては、地域の目標のうち公共交通が貢献すべき領域を設定する、政策目標設定から関わる形態が望ましい。一方で交通サービスに関しては、他行政サービスと同列に比較され、その地域への貢献によって選択されるべきであろう。

現状の市民参加は、あくまで目の前にある課題の解決を旨としており、実証主義に基づいた主張通りに情報提供と労力提供の役割を持っている。しかしそれだけでは、全ての計画段階における批判はなし得ないものであり、少なくとも交通の「要件定義」段階である目標立案レベルへの市民参加は、公共サービスの一方性を排するために重要な要素である。

こうした目標への市民参加は、それ自体がハンス・ケルゼンの指摘する民主主義の本質と深く結びついている。政策の目標、手段、あらゆる点に対して市民に批判するための回路を残すことで、武力によるクーデターがなくとも政権の構造転換をなしえる、というのがその主張である。

3.5 実装ギャップの実証分析：Japan MaaS

3.5.1 分析手法

分析手法について、「計画における計画指標（Key Performance Indicator, 以下 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。第1の分析項目については、MaaS が「事業の手段」としてではなく、「政策の手段」として機能しているかを確認するものである。この点から、評価指標の状況について観察対象とした。MaaS 事業自体の事業性や交通事業の利用状況だけではなく、それ以外の評価指標を計画に置いているかを確認する。

第2の分析項目については、地域における交通事業者など公益的な企業の特性を鑑み、市民参加が行われており目標や手段への批判プロセスがあるかを確認する。市民の行政への批判回路について、議会を通じた参加等はあるが、交通計画では協議会参加、グループヒアリング、アンケート調査、パブリックコメントが推奨手法として取り上げられている。これらのうち、合議への参加を通じて直接の批判とできるのは協議会参加のみである。

以上の検討より、会議体への市民の参加を指標としてとらえることとした。

3.5.2 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。

ここで、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について確認し、類型化を試みる。日本版 MaaS の主要な要素技術として、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。なお、参入の高度化に必要な API をはじめとしたデータ連携のオープン化に関して言及しているのは 2 事例のみであった（加賀、しんゆり）。

3.5.3 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 1 の分析項目について、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。さらに第 2 の分析項目については、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 3.1: Japan MaaS 36 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%

3.5.4 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。

3.5.5 考察

分析対象とした日本版 MaaS 推進事業へ認定されたそれぞれは、その目標を考えると、ほとんどの場合交通事業の成立、また日本版 MaaS 事業の成立を目指している。実際のところ、本事業を担当する国土交通省新モビリティサービス推進課の担当者は、「申請前の 1 年で事業検討して、申請後の 1 年で事業として自走できるようにしてほしい」と述べており、事業としての成立を企図している。一方で、こうしたスタンスは実際のところ、海外ではすでに行き詰まり感を見せている。日本における MaaS は、こうした国際的な流れを反映していない。

また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べきである。交通事業、MaaS事業の指標のみでは、こうした運営には不十分である。それは交通事業の、利益を産む/費用を抑える事業としての価値こそ管理できるが、地域交通事業はこうした効率化が難しい現状を鑑みると、地域における機能については管理できないからである。

さらに、こうした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くと推察される。

3.5.6 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。

このような状況になった要因として、いくつかの仮説を挙げておきたい。第 1 に、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があげられる。一般に、デジタル時代の政府活動への意向にはそれに即した評価、批判のシステムへ移行する必要があるが、それには取引費用がかかりすぎることが分かっている。第 2 に、これは第 1 の指摘と関連するが、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があげられる。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は十分にある。

3.6 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、いくつかの理由から本研究の「実験場」として適している。第一に、複雑なステークホルダー関係である。国、自治体、交通事業者、市民など多様な主体が関与している。第二に、明確な政策目標である。持続可能な公共交通の維持・発展という目標が明確である。第三に、実装ギャップの可視化である。制度的意図と実践の乖離が観察可能である。第四に、データの入手可能性である。政策文書、統計データへのアクセスが比較的容易である。

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」が謳われていても、実践レベルでは事業効率性の追求に偏向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くと推察される。

「利用者視点」を導入しているう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

4.1.1 研究の背景

日本の地域公共交通政策は、2002年の規制緩和以来、「連携」(collaboration)と「共創」(co-creation)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、より広範な公共ガバナンスの潮流を反映している。

この変革の制度化は、近年の政策展開によってさらに加速している。地域公共交通計画の策定は義務的努力となり、ステークホルダー間の効率的な情報交換と協議を促進する会議体の設立を強調する具体的なガイダンスが示されている。この制度的枠組みは、連携と共創のプロセスが機能すべき正式な構造を提供している。

4.1.2 政策の文脈と変遷

日本の地域公共交通政策は、市場メカニズムと公的介入のバランスに対する異なるアプローチを反映した、明確な段階を経て進化してきた：

2002年規制緩和 路線運行を許可制から届出制に移行し、補助金配分を会社全体から路線別支援に変更し、ターゲットを絞った補助金を通じて本質的なサービスを維持しながら市場競争を導入した。

2006年地域公共交通会議 正式的な協調ガバナンスのメカニズムを確立し、コミュニティバスの運行を可能にし、交通計画へのステークホルダー参加のプラットフォームを提供した。

2010年生活交通サバイバル事業 「努力する地域」のみが生活交通サービスを維持することを許可する競争的アプローチを制度化し、地方のイニシアティブと自立を強調した。

2021年共創フレームワーク 共創を中心的組織原則として明示的に採用し、リソース動員（「交通リソースの総動員」）、共同管理（事業者間の直接協働）、サービス最適化（「束ねて削減」アプローチ）を包含した。

4.1.3 実装ギャップの実証的証拠

連携と共創の原則の実践的な実装を評価するために、共創政策の具体的な現れである日本の MaaS (Mobility as a Service) イニシアティブの実証分析を行った。2020 年度 Japan MaaS イニシアティブで認定された全 38 プロジェクトの分析は、政策の理想と実装の現実の間に大きな乖離を明らかにした。

事業パフォーマンス指標に関して、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。

非事業指標に関して、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。

市民参加メカニズムに関して、最も懸念されるのは、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意味ある市民参加のメカニズムを確立していたに過ぎないことである。

Table 4.1: Japan MaaS 38 プロジェクトの実装ギャップ分析

指標タイプ	プロジェクト数	割合
事業パフォーマンス指標	35/38	92%
非事業指標（社会的影響等）	11/38	29%
市民参加メカニズム	2/38	5%

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。

4.1.4 研究目的とアプローチ

本研究は3つの主要な研究問い合わせ取り組む。第一に、連携的交通政策がその表明された目的をどの程度達成しているかを検討すること。第二に、認知バイアスが協調的調整の効果にどのように影響するかを計算論的モデリングアプローチを用いて探求すること。第三に、連携の利益を最適化しながらその限界を緩和するのに役立つ制度的配置について議論することである。

本研究は、実証的政策評価、協調ロボット制御理論を用いた計算論的分析、そして両方の流れから情報を得た制度設計についての議論を組み合わせた統合的アプローチを採用する。本研究は本質的に探索的であり、計算論的モデルは認知バイアス効果に関する仮説を生成する類推的枠組みとして機能し、さらなる実証的検証を必要とする。

4.2 理論的枠組み

4.2.1 協調ガバナンス理論

協調ガバナンスは公共行政において支配的なパラダイムとして登場し、Ansell and Gash (2008) によって「1つ以上の公共機関が、非政府ステークホルダーを、形式的で、合意志向で、審議的な集団的意思決定プロセスに直接関与させ、公共政策を立案または実施し、または公共プログラムまたは資産を管理することを目的とする統治配置」と定義されている。

交通政策の文脈において、協調的アプローチは伝統的な規制フレームワークの限界に対する解決策として推進されてきた。Kato et al. (2009) は、成功する地域住民参加型公共交通のための4つの重要な条件を特定している：ステークホルダー間の共有認識と責任分担、すべての参加者にとっての相互利益、主要な調整者の存在、そしてステークホルダーの努力とサービス改善成果の間のつながりである。

しかし、協調ガバナンスは固有の課題に直面している。Emerson et al. (2012) は、協調には多大な取引コストが必要であり、最低共通分母の解につながる可能性があり、組織化された利益によって捕捉される可能性があると指摘している。これらの課題は、技術的複雑さ、規制上の制約、商業的利益が追加的な調整の困難を生み出す交通政策において特に深刻である。

4.2.2 公共サービス提供における共創

共創は伝統的な協調を超える進化を表し、公共機関とステークホルダー間の共同価値創造を強調する (Voorberg et al., 2015)。交通政策において、共創はいくつかの形態で現れる：非伝統的な交通リソースを活用するリソース動員、事業者間の正式なパートナーシップを通じた共同運営、そ

して他のサービスと交通を束ねて包括的なモビリティソリューションを創出するサービス統合である。

九州運輸局（2023）は、共創が以前は不可能だった路線再編を可能にしサービス効率を改善する一方で、事業パフォーマンスの改善は限定的であり、必要な信頼関係の構築にはかなりの時間投資が必要であることを発見した。

4.2.3 協調的政策実装における認知バイアス

協調的交通政策実装に大きな影響を与えるいくつかの認知バイアスがある。

現状維持バイアス (Status quo bias) は、現在の条件を維持することへの体系的な好みを反映し、革新的な交通政策の実装を困難にする（Watkins & Musselwhite, 2025; Samuelson & Zeckhauser, 1988）。

確証バイアス (Confirmation bias) は、ステークホルダーが既存の信念を確認するような方法で情報を探したり解釈したりするときに現れ、代替的な視点への開放性を低下させる（Rastogi et al., 2020; Acciarini et al., 2020）。

狭い視野 (Narrow framing bias) は、ステークホルダーがシステム全体の最適化よりも局所的な最適化に焦点を当てるときに発生する（Kahneman, 2011）。

これらのバイアスは、政策設定において異なるメカニズムを通じて作用する。それらは情報処理に影響を与え（ステークホルダーがデータをどのように知覚し優先順位を付けるかに影響することによって）、ステークホルダー間の相互作用を複雑にし（マルチステークホルダー設定における認知的コンセンサスの達成を妨げることによって）、そして意思決定のダイナミクスを形成する（政策革新を遅らせたりブロックしたりすることによって）。

重要なことに、認知バイアスは有害な効果と潜在的に建設的な効果の両方を示すことができる。ほとんどのバイアスは調整を妨げるが、構造化されたステークホルダーエンゲージメントと知識交換は、根本的な仮定を明らかにし、時には優先順位をシフトさせ意思決定の質を向上させることができる（Knoppen et al., 2021; Watkins & Musselwhite, 2025）。

4.2.4 協調制御理論とマルチエージェント調整

協調制御理論は、複数の自律エージェントが共有目標を達成するために調整するための枠組みを提供する。この理論は、コンセンサスメカニズム、フォーメーション問題、およびリソース配分を、分散的かつ適応的な方法を通じて扱う（Gulzar et al., 2018）。社会システムへのその拡張は、複雑で分散化された環境を管理するために、分散的意思決定と適応的調整の原則を活用する。

本研究は、Yoshihara et al. (2009) の協調ロボット制御フレームワークを適用する。これは、N 関節ロボットアームが局所情報交換を通じてどのように協調動作を達成するかをモデル化するものである。類推的マッピングを可能にする重要な洞察は、ロボットアームの関節と政策ステークホルダーの両方が、局所情報制約と個別の行動傾向の下で運用しながら、共有目標に向けて行動を調整しなければならないということである。

4.3 計算論的モデリング・フレームワーク

4.3.1 類推的マッピング：正当化と限界

本研究は、N 関節ロボットアームを階層的政策ネットワークの類推的モデルとして使用する。このマッピングでは、各関節 i は特定のガバナンスレベルでの政策ステークホルダーに対応し、システムの集合的挙動は政策ネットワークの調整ダイナミクスを表現する。

直列チェーントポロジーの正当化：日本の交通ガバナンスは階層構造を通じて運用される。国が政策フレームワークを設定し、都道府県が地域計画を調整し、市町村が地域交通施策を実施し、事業者がサービスを提供する。このガバナンス階層は直列依存性を示す—市町村レベルの決定は都道府県および国のフレームワークによって実質的に制約され、運用上の決定は市町村計画によって制約される。直列運動連鎖は、各レベルの「位置」（政策成果）が先行レベルの累積決定に依存するこの階層的依存構造を捉えている。

認められた限界：この類推的マッピングのいくつかの重要な限界を認める必要がある。トポロジーについては、実際のステークホルダーネットワークは純粋に直列ではなく、並列およびメッシュの相互作用を示す。直列チェーンは階層的依存性を捉えるが、同じガバナンスレベルでのステークホルダー間の水平的調整は捉えない。固定リンク長については、ステークホルダーの影響力 (a_i) は定数として扱われるが、実際には影響力は動的で文脈依存である。この単純化は、調整期間にわたって安定した制度的役割を仮定している。2次元出力空間については、モデルは \mathbb{R}^2 で動作し、これは2つの主要な政策次元（例えば効率性とアクセシビリティ）を表現すると解釈できる。実際の政策成果はより高次元であり、この削減は2つの優先された目標間のトレードオフを捉えている。単一ターゲットについては、モデルは共有政策目標 x_d を仮定する。ステークホルダーはしばしば対立する目標を持つが、この単純化は、合意ターゲットが民主的プロセスを通じて確立されたシナリオをモデル化する。

4.3.2 モデル構造

各関節 i は状態変数を持つ：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 $\theta_i \in \mathbb{R}$ 、およびリンク長 $a_i \in \mathbb{R}^+$ 。

Table 4.2: シンボル-概念対応表

記号	数学的意味	単位	制度的意味	理論的根拠
θ_i	関節角度	rad	ステークホルダーの政策立場	Yoshihara et al.
a_i	リンク長	m	ステークホルダーの制度的影響力(固定)	Yoshihara et al.
$b_{sq,i}$	現状維持バイアス係数	–	変化への抵抗	Samuelson & Zeckhauser; bounded confidence
$b_{cf,i}$	確証バイアス係数	–	自己判断への過信	DeGroot モデルの非対称重み
$b_{nf,i}$	狭い視野係数	–	局所最適化への焦点	限定合理性モデル
k_i	協調係数	–	他者との協調度	合意形成モデル

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。

速度は位置の時間微分として計算される：

$$\mathbf{v}_i = \frac{d\mathbf{x}_i}{dt} \quad (4.2)$$

4.3.4 制御目標と調整メカニズム

システムの目標は、政策目標を表現する目標位置 \mathbf{x}_d を追跡するようにエンドエフェクタ（最終ステークホルダー）を導くことである。各関節 i は協調制御アルゴリズムに従う。

目標方向ベクトル計算：目標からエンドエフェクタへの方向ベクトル：

$$\tilde{\mathbf{v}}_i = \mathbf{x}_d - \mathbf{x}_N \quad (4.3)$$

協調係数による重み付け：各関節は協調係数 k_i に基づいて調整に参加する：

$$\mathbf{v}_{target,i} = k_i \cdot \tilde{\mathbf{v}}_i \quad (4.4)$$

4.4 認知バイアスの統合

4.4.1 現状維持バイアスのモデル化

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。Friedkin-Johnsen モデルのアンカリングパラメータに基づき、より小さい感受性値はより大きな変化への抵抗をもたらす：

$$\frac{d\theta_i}{dt} = (1 - b_{sq,i}) \cdot \omega_i \quad (4.5)$$

ここで ω_i はバイアスなしの角速度である。 $b_{sq,i} = 0$ は完全な感受性を表し、 $b_{sq,i} = 1$ は完全な固定を表す。

4.4.2 確証バイアスのモデル化

確証バイアス $b_{cf,i} \in [-1, 1]$ は、DeGroot スタイルの意見ダイナミクスにおける類似性依存重み付けに基づき、協調係数の修正として組み込まれる：

$$k'_i = k_i \cdot (1 + b_{cf,i} \cdot 0.5) \quad (4.6)$$

正の値は自己判断への過信を表し、他者からの影響を減らす。負の値は他者の判断への過度の感受性を表す。

4.4.3 狹い視野のモデル化

狭い視野 $b_{nf,i} \in [0, 1]$ は、限定合理性モデルに基づき、全体目標ベクトルの削減として組み込まれる。エージェントがグローバルパフォーマンス指標ではなく局所報酬信号に制限される状況をモデル化する：

$$\tilde{\mathbf{v}}'_i = (1 - b_{nf,i}) \cdot \tilde{\mathbf{v}}_i \quad (4.7)$$

$b_{nf,i} = 0$ は完全なグローバル視野を表し、 $b_{nf,i} = 1$ は完全な局所視野（目標ベクトルがゼロになる）を表す。

4.5 シミュレーション実験

4.5.1 実験設計

8 関節ロボットアームを用い、日本の階層的ガバナンス構造（国→都道府県→市町村→事業者→...）をモデル化した。1,820 回のシミュレーション実験をランダム化された初期条件で実施した。

実験パラメータ：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、リンク長 a_i ：0.5 から 2.0 の間でランダムに設定、各バイアス値：0.0 から 1.0 の範囲で段階的に変化させた。

4.5.2 評価指標

以下の評価指標を用いた：

精度 (A_t) 最終エンドエフェクタ位置と目標位置の間のユークリッド距離

エネルギー効率 (E_t) シミュレーション期間中の総関節動作

関節活動度 ($J_{act,i}$) 各関節の角度変化の合計

4.6 結果

4.6.1 確証バイアスの効果

確証バイアスは、統計的に有意な調整低下を引き起こさなかった ($F = 0.838, p = 0.602$)。これは直感に反する結果に見えるかもしれないが、構造化された協調設定では、一定レベルの自己確信が実際には意思決定プロセスを加速し、集団パフォーマンスを損なうことなく集団コンセンサスへの到達を容易にする可能性があることを示唆している。

重要なことに、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。

4.6.2 現状維持バイアスの閾値効果

現状維持バイアスは、統計的に有意な負の効果をもたらした ($F = 1.593, p = 0.044; \eta^2 = 0.039$)。より詳細な分析により、顕著な**閾値効果**が明らかになった。

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。

Table 4.3: 現状維持バイアスの閾値効果

バイアス値	精度低下	解釈
0.0–0.25	5%以下	順調な調整
0.25–0.50	15–30%	調整困難
0.50–1.00	50%以上	調整失敗

4.6.3 狹い視野の一貫した負の影響

狭い視野は、有意な線形劣化を示した ($F = 1.985, p = 0.028; R^2 = 0.204$)。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。これは、各ステークホルダーが自分の局所的な最適化を追求し、システム全体のパフォーマンスを考慮しないためである。

4.7 統計的サマリー

4.8 制度的設計への示唆

計算論的分析と実装ギャップの実証分析に基づき、本研究は 3 層の制度的アーキテクチャを提案する：

Table 4.4: 認知バイアス効果の統計的サマリー

バイアス種別	F 値	p 値	効果 size	推奨対処	
確証バイアス	0.838	0.602	—	適度な自信の許容	* $p < 0.05$
現状維持バイアス	1.593	0.044*	$\eta^2 = 0.039$	段階的導入	
狭い視野	1.985	0.028*	$R^2 = 0.204$	全体目標の可視化	

層 1：民主的品質設定 政治-行政インターフェースにおいて、政策目標と品質基準を民主的に設定する。この層では、合意ターゲット x_d が確立される。

層 2：行政的調整 行政-事業インターフェースにおいて、ステークホルダー間の調整を管理する。現状維持バイアスの閾値を考慮した段階的導入を実施する。

層 3：運用実装 事業レベルにおいて、サービスを提供する。狭い視野を防ぐため全体目標の可視化を行う。

この 3 層アーキテクチャは、民主的アカウンタビリティと運用効率の両方を最適化するために、政治-行政インターフェースと行政-事業インターフェースを分離する。

4.9 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。第一に、確証バイアスは統計的に有意な調整低下を引き起さなかった ($F = 0.838, p = 0.602$)。適度な範囲では逆説的に協調を促進する可能性がある。第二に、現状維持バイアスは統計的に有意な負の効果をもたらした ($F = 1.593, p = 0.044$)。閾値効果があり、約 0.25 を超えると臨界点を超えて急激に協調が阻害される。第三に、狭い視野は有意な線形劣化を示した ($F = 1.985, p = 0.028$)。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。特に、現状維持バイアスの閾値効果は、政策変革を段階的に導入する必要性を示唆しており、狭い視野の一貫した負の影響は、全体目標の可視化と共有の重要性を強調している。

次章では、生成 AI と人間の関係性について、ZK-SNARKs 型政策評価システムを通じて考察する。

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。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、第2章で述べたルーマンの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。

本章では、こうした課題に対処するため、ZK-SNARKs (Zero-Knowledge Succinct Non-interactive Arguments of Knowledge) の概念を援用した政策評価システムを提案する。

5.2 ZK-SNARKs 概念の援用

5.2.1 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

5.2.2 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。第一に、処理速度の革新である。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した。第二に、証明サイズの一定性である。100 万ステップの複雑な計算であっても証明サイズは 288 バイト程度と一定である。第三に、適用範囲の広さである。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である。第四に、信頼性の高さである。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である。

5.2.3 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、いくつか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する場面、個人が個人情報を守りながら、専門性を証明する場面、そして行政が内部情報を守りながら、政策判断の根拠を説明する場面などがその例である。

5.3 ZK-SNARKs 型政策評価フレームワークの提案

5.3.1 フレームワークの目的

本研究では、以下の 4 つの目的を持つ ZK-SNARKs 型政策評価フレームワークを提案する：

第一に、プライバシー保護型知識共有の実現：関係者が機密性の高い専門知識や経験の詳細を開示することなく貢献できる仕組みを構築し、企業秘密や個人情報を保護しながら社会課題解決に必要な知識を活用する。

第二に、アルゴリズムによる信頼性の技術的保証：従来の主観的信頼関係に依存しないアルゴリズムによる情報の確からしさ検証により、専門知識の妥当性をアルゴリズム的に評価し、権威主義的な意思決定を回避する。

第三に、効率的な合意形成プロセス：複雑な検証手続きの自動化による意思決定の迅速化と、大規模かつ多様なステークホルダーチームとの効率的な協議の実現。

第四に、透明性と秘匿性の両立：意思決定プロセスの透明性を保ちながら個別の機密情報を保護し、政策根拠の存在証明による行政説明責任を強化する。

これらの目的により、従来不可能であった「秘密を明かすことなく専門性の確からしさを提供する」政策対話の新たなパラダイムを創出することを目指している。

5.3.2 入札型秘匿証明の構築

ZK-SNARKs 型の仕組みにおいては、既存の評価に関する議論が適用できる。この枠組みとして、入札のような「中身を開示しないけれど目的達成はわかる」という仕組みをその都度構築することが求められる。

具体的には、各政策課題や意思決定局面において、関係者が以下の 4 つの証明要件を満たす証明システムを動的に設計する必要がある：

能力証明 課題解決に必要な専門性や資源を保有していること（詳細な手法や情報源は秘匿）

実績証明 類似課題での成功事例や経験の存在（具体的な顧客名や機密情報は秘匿）

コミット証明 提案する解決策の実現可能性と期待成果（実装詳細や競争優位情報は秘匿）

品質保証 提供する成果物の品質基準達成（内部プロセスや技術仕様は秘匿）

このようなシステムは、従来の公開入札制度が持つ透明性の利点を維持しながら、関係者のプライバシーと競争優位性を保護し、より多様で質の高い提案を引き出すことを可能にする。

5.4 LLM as a Judgeによる実装

5.4.1 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単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単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。

5.4.2 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。

5.4.3 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。

政策適用における特有の課題：政策判断の複雑性と多面性への対応限界、社会的価値観や倫理的考慮の処理困難、ステークホルダーの多様性への配慮不足が挙げられる。

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 LLM as a Judge の限界を超えるための人間支援機能

5.5.1 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。第一に、憲法的ルールセットである。人間が事前に定義した行動規範・価値基準のリストを用意する。第二に、自己批判・自己修正機能である。AI が自身の回答を憲法に照らして評価し改善する。第三に、最小限の人間監督である。詳細な判定ではなく原則設計のみに人間が関与する。

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、主観的判断の客觀化が実現される。

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こと、民主的正当性を確保すること、そして透明性と説明責任を満たすことが可能となる。

5.5.2 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み解き・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。

5.5.3 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。具体的には、AI による一次評価を行い、評価結果に対する異議申立てを受付、人間による二次評価を行い、最終判断を提示するというプロセスである。

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。

5.5.4 決定論的運用

LLM の確率的な出力を制御するため、いくつかの手法を組み合わせる。Temperature=0 によりランダム性を排除し、Self-Consistency により複数回の出力から一貫性のある結果を選択し、TEE (秘匿実行環境) により処理の透明性を確保する。

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 値値統合、評価基準の設計

このアーキテクチャにより、Constitutional AI 原則に基づく評価基準の設計、計画・推論分離型の EvalPlanner システム、控訴プロセス型の段階的人間介入、そして複数 LLM によるメタジャッジフレームワークを組み合わせることで、効率性と信頼性を両立する ZK-SNARKs 型政策評価システムの実現が可能となる。

5.6.2 生成 AI の「杖」としての位置づけ

このシステムにおける生成 AI は、第 2 章で述べた「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、いくつかの原則に基づく。第一に、AI は人間の最終判断を前提とする。第二に、AI の限界を明示的に理解する。第三に、人間-AI 協調のプロセスを透明化する。第四に、AI 自体のバイアスに対処する。

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。したがって、価値判断、合意形成、説明責任は人間が担うべきである。

5.6.3 限界と課題

このシステムにはいくつかの限界がある。第一に、数学的保証と確率的期待の違いである。ZK-SNARKs の数学的完全性は、LLM では実現できない。第二に、AI 自体のバイアスである。学習データに含まれるバイアスが評価結果に影響する。第三に、透明性の限界である。LLM の内部処理の完全な説明は困難である。

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。

まず、ウィキッド・プロブレム時代における政策評価の課題を整理し、コミュニケーションの二重の困難（理解の多様性、知識の非対称性）を指摘した。

次に、ZK-SNARKs の概念を援用し、「秘密を守りながら専門性を証明する」仕組みとしての政策評価フレームワークを提案した。このフレームワークは、プライバシー保護型知識共有、アルゴリズムによる信頼性保証、効率的な合意形成、透明性と秘匿性の両立という 4 つの目的を持つ。

さらに、LLM as a Judge による実装の可能性と限界を論じ、Constitutional AI、計画・推論分離型システム、控訴プロセス、決定論的運用を通じて、人間がアルゴリズムの改善者として機能する新しいパラダイムを示した。

最後に、三層アーキテクチャ（ZK-SNARKs 秘匿証明層、LLM as a Judge 評価層、Constitutional AI + 市民討議層）を提案し、生成 AI を人間の「執政の創造性」を補完する「杖」として位置づけた。このシステムは、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し、秘匿性と信頼性を両立する評価プロセスを提供し、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するものである。

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としては、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ことが重要である。確証バイアスの活用としては、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ことが有効である。狭い視野への対処としては、全体目標の可視化と横断的指標の導入を必須とすることが求められる。

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。第一に、秘匿性の制度化である。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が必要である。第二に、自動評価の境界である。AIによる評価と人間による判断の境界を明確化する必要がある。第三に、市民参加の質的向上である。市民討議を通じた評価基準の共同設計が求められる。

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、小さな変化の積み重ねという戦略を提案する。具体的には、パイロット導入として小規模な実験から開始し、成功事例を蓄積することが有効である。また、段階的拡大として、臨界点を超えないよう、徐々に適用範囲を拡大することが求められる。さらに、デフォルト設定の変更として、現状維持の方向に働く制度上のデフォルトを見直すことも重要である。

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、多様な視点の構造的導入という戦略を提案する。「悪魔の代理人」の制度化として、AIまたは人間による批判的視点の提示を必須化することが有効である。また、多様なステークホルダーの参画として、異なる立場・利害を持つ主体の関与を確保することが求められる。さらに、反証可能性の確保として、仮説を反証する証拠を積極的に探索するプロセスを組み込むことも重要である。

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、全体目標の可視化という戦略を提案する。横断的評価指標として、単一の指標ではなく、複数の指標による評価を導入することが有効である。また、システムマップ

の作成として、政策の全体像を可視化した図の共有が求められる。さらに、長期的影響の分析として、短期的な成果だけでなく、長期的な影響も評価することが重要である。

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。政治的判断については、人間（政治家・議会）が最終的に決定する。行政的分析については、AIが情報の整理・分析を支援する。民主的アカウンタビリティについては、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。政策の具体化については、AIが選択肢の生成を支援する。運用の効率化については、AIが日常的な判断を補助する。人間による監督については、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。第一に、評価基準の共同設計である。市民討議を通じた基準の策定が必要である。第二に、透明性の確保である。AIの評価プロセスと基準の公開が求められる。第三に、控訴の権利である。評価結果に対する異議申立ての機会を確保する必要がある。第四に、人間による最終判断である。AIの評価は参考情報として位置づけられる。

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。

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

S 層では現状維持バイアスによる変化抵抗が課題となるため、長期ビジョンの共有と段階的目標設定が有効である。T1 層では狭い視野による部分最適化が課題となるため、横断的指標と全体目標の可視化が求められる。T2 層では確証バイアスによる視点の固定化が課題となるため、多様な視点の構造的導入が必要である。O 層では日常的な認知バイアスが課題となるため、AIによる対話的介入が有効である。

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。第一に、補完性の原則である。AIは人間の能力を補完し、代替しない。第二に、透明性の原則である。AIの活用方法と限界を明示する。第三に、アカウンタビリティの原則である。最終的な判断と責任は人間が負う。第四に、継続的学习の原則である。AIと人間の協調プロセスを継続的に改善する。

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。主な貢献として、認知バイアスへの具体的な対処戦略の提示、生成AIを組み込んだ制度設計の方向性の提示、「連携・共創」の再設計に向けた枠組みの提示が挙げられる。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。第一に、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こと。第二に、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こと。第三に、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と。

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效果、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として、段階的変化導入の設計を提案した。確証バイアスへの対処として、「悪魔の代理人」の制度化を提案した。狭い視野への対処として、横断的評価指標の導入を提案した。生成 AI の活用として、補完性、透明性、アカウンタビリティの原則を提示した。

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がその主要な内容である。

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）、長期的影響（数十年単位での都市構造の変化）という点で興味深い研究対象となる。

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、いくつかの課題がある。技術的実装としては、TEE、LLM、Constitutional AI の統合が必要である。制度的設計としては、法的位置づけや運用ルールの策定が求められる。社会的受容としては、市民の理解と信頼の獲得が重要である。

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] Georges Canguilhem. *Le normal et le pathologique*. 滝沢武久訳：正常と病理（叢書・ウニベルシタス, 225）、法政大学出版局、1987年. PUF, 1966.
- [3] Takashi Iba. “An Autopoietic Systems Theory for Creativity”. In: *The 1st Conference on Collaborative Innovation Networks (COINs)*. 2009.
- [4] John W. Kingdon. *Agendas, Alternatives, and Public Policies*. Little, Brown, 1984.
- [5] 草野厚. **政策過程分析入門**. 慶應義塾大学出版会, 1997.
- [6] Paul A. Sabatier. “An Advocacy Coalition Framework of Policy Change and the Role of Policy-Oriented Learning Therein”. In: *Policy Sciences* 21.2/3 (1988), pp. 129–168.
- [7] 村松岐夫. **戦後日本の官僚制**. 東洋経済新報社, 1981.
- [8] Horst W. J. Rittel and Melvin M. Webber. “Dilemmas in a General Theory of Planning”. In: *Policy Sciences* 4.2 (1973), pp. 155–169.
- [9] 杉谷和哉. “**ウィキッド・プロブレムとしての新型コロナ感染症 政治と専門性の関係を中心に**”. In: **医療福祉政策研究** 4.1 (2021), pp. 27–37.
- [10] Jürgen Habermas. *Theorie des kommunikativen Handelns*. 河上倫逸ほか訳：コミュニケーション的行為の理論、未来社、1985–1987年. Suhrkamp, 1981.
- [11] Niklas Luhmann. *Soziale Systeme*. 馬場靖雄訳：社会システム理論、勁草書房、2001–2020年. Suhrkamp, 1984.
- [12] Georg Kneer and Armin Nassehi. *Niklas Luhmanns Theorie sozialer Systeme*. 館野受男訳：ルーマン社会システム理論——「知」の扉をひらく、新泉社、1995年. Fink, 1993.
- [13] Humberto R. Maturana and Francisco J. Varela. *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel, 1980.
- [14] Ben Shneiderman. *Human-Centered AI*. Oxford University Press, 2022.
- [15] 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.
- [16] 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.
- [17] Hirokazu Kato et al. “Community-Participatory Regional Public Transport”. In: *Transport Policy Studies* 12.2 (2009), pp. 2–9.
- [18] 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.
- [19] William Samuelson and Richard Zeckhauser. *Status Quo Bias in Decision Making*. Vol. 1. 1. 1988, pp. 7–59.
- [20] Federica Russio et al. “Confirmation Bias in Policy Decision Making”. In: *Policy Sciences* 48.3 (2015), pp. 289–313.

- [21] 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.

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