

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：〇〇 教授

〇〇大学 大学院 〇〇研究科  
〇〇専攻 博士後期課程

2026 年 〇月 〇日 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）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	x
<b>1 序論</b>	<b>1</b>
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	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
<b>2 先行研究のレビュー：Human-AI Policy の議論</b>	<b>3</b>
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

<b>3</b>	<b>舞台としての公共交通政策：現状と課題</b>	<b>7</b>
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	Japan MaaS の概要	8
3.4.2	分析結果	8
3.4.3	構造的要因	9
3.5	小括：公共交通政策を「実験場」として位置づける理由	9
<b>4</b>	<b>認知バイアスの政策協調への影響：計算論的分析</b>	<b>10</b>
4.1	研究設計	10
4.1.1	協調ロボット制御モデルによる政策ネットワーク表現	10
4.1.2	三つの認知バイアスのモデル化	10
4.2	シミュレーション実験	11
4.2.1	実験条件	11
4.2.2	評価指標	11
4.3	結果と考察	11
4.3.1	現状維持バイアスの閾値効果	11
4.3.2	確証バイアスの逆説的效果	11
4.3.3	狭い視野の一貫した負の影響	11
4.4	小括：認知バイアスによる協調失敗のメカニズム	11
<b>5</b>	<b>生成 AI と人間の関係性：ZK-SNARKs 型政策評価システム</b>	<b>12</b>
5.1	生成 AI の位置づけ	12
5.1.1	人間の「執政の創造性」を補完する「杖」	12
5.1.2	AI ができないこと	12
5.2	認知バイアスへの対話的介入	12
5.2.1	「悪魔の代理人」機能	12
5.2.2	視野拡大の支援	13
5.2.3	メタ認知の促進	13
5.3	ZK-SNARKs 概念の援用	13
5.3.1	秘密を守りながら専門性を証明する	13
5.3.2	ZK-SNARKs の 4 つの特性の政策評価への翻訳	13
5.4	LLM as a Judge による実装	13
5.4.1	Constitutional AI と市民討議	13
5.4.2	決定論的運用	14
5.4.3	控訴プロセスによる人間介入	14
5.5	ZK-SNARKs 型政策評価のアーキテクチャ	14
5.5.1	三層アーキテクチャ	14
5.5.2	限界と課題	14
5.6	小括	14
<b>6</b>	<b>制度設計への示唆</b>	<b>15</b>
6.1	理論的含意	15
6.1.1	第 4 章のシミュレーション結果からの設計原則	15
6.1.2	第 5 章の ZK-SNARKs システムからの制度的含意	15
6.2	各バイアスへの対処戦略	15

6.2.1	現状維持バイアス：小さな変化の積み重ね	15
6.2.2	確証バイアス：多様な視点の構造的導入	15
6.2.3	狭い視野：全体目標の可視化	16
6.3	生成 AI を組み込んだ制度設計	16
6.3.1	政治的-行政的インターフェース	16
6.3.2	行政的-事業的インターフェース	16
6.3.3	ZK-SNARKs 型システムの制度的位置づけ	16
6.4	「連携・共創」の再設計	16
6.4.1	三層制度設計（S-T1-T2-O）の再考	16
6.4.2	生成 AI を「杖」として活用するガバナンス	16
6.5	小括	17
<b>7</b>	<b>結論：都市計画への展開</b>	<b>18</b>
7.1	研究の総括	18
7.1.1	研究目的の達成	18
7.1.2	核となる主張	18
7.2	理論的貢献	18
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	18
7.2.2	ZK-SNARKs 概念の政策評価への応用	19
7.2.3	生成 AI の「杖」としての理論的位置づけ	19
7.3	実践的貢献	19
7.3.1	制度設計への提言	19
7.3.2	ZK-SNARKs 型政策評価システムの設計指針	19
7.4	今後の課題：都市計画を舞台にした実証	19
7.4.1	より複雑な政策領域への適用	19
7.4.2	ZK-SNARKs システムの社会実装	19
7.4.3	生成 AI と人間の協調的関係性の継続的検証	20
7.5	結び	20
<b>A</b>	<b>付録</b>	<b>22</b>
A.1	協調ロボット制御モデルの数式展開	22
A.1.1	運動学モデル	22
A.1.2	協調係数の計算	22
A.2	Japan MaaS プロジェクト分析の詳細	22
A.2.1	分析対象プロジェクト一覧	22
A.2.2	評価指標のコーディング基準	22
A.3	ZK-SNARKs 型政策評価システムの実装詳細	23
A.3.1	システム構成	23
A.3.2	技術的仕様	23



# List of Figures

# List of Tables

3.1	STO フレームワークによる政策分析 . . . . .	7
3.2	Japan MaaS 38 プロジェクトの評価指標分析 . . . . .	8
5.1	ZK-SNARKs 特性の政策評価への翻訳 . . . . .	13
6.1	認知バイアスと生成 AI を考慮した三層制度設計 . . . . .	17
A.1	評価指標のコーディング基準 . . . . .	23
A.2	技術的仕様 . . . . .	23

# 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

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」<sup>1</sup>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<sup>2</sup>によれば、92%のプロジェクトが事業指標のみを重視し、社会的影響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

---

<sup>1</sup> 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

<sup>2</sup> 第 3 章で詳述

- **確認バイアス (Confirmation Bias)**：自己の信念を確認する情報の優先的選択
- **狭い視野 (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**：人間的な自発的な揺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 と人間の協調的關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 と人間の關係性を理論的・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 と人間の關係性：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の中心となっている [2]。

#### 2.2.2 「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

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

### 2.2.3 AIの「杖」としての位置づけ

生成AIを「杖 (Aarons' 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・増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、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

## 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

### 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

### 2.6.2 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。

### 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（雲南市だんだんタクシー）

##### 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 Japan MaaS の概要

Japan MaaS は、MaaS（Mobility as a Service）の概念を日本の地域課題解決に応用する公的政策イニシアティブである。2020 年度に認定された 38 プロジェクトを対象に、計画書の内容分析を行った。

#### 3.4.2 分析結果

分析の結果、以下の実装ギャップが明らかになった：

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	35	92%
非事業指標を含む	11	29%
市民参加の仕組み	2	5%

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。

### 3.4.3 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）

### 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

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 研究設計

#### 4.1.1 協調ロボット制御モデルによる政策ネットワーク表現

本研究では、Yoshihara et al. (2009) [3] の協調ロボット制御理論を応用し、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、政策ネットワークとして以下のように解釈される：

- 各関節  $i$  は政策ステークホルダーに対応
- 関節角度  $\theta_i$  はステークホルダーの政策立場を表す
- リンク長  $a_i$  はステークホルダーの影響力・能力を表す
- エンドエフェクタ位置は政策目標の達成状態を表す

#### 4.1.2 三つの認知バイアスのモデル化

##### 現状維持バイアス (Status Quo Bias)

現状維持バイアス  $b_{sq,i} \in [0, 1]$  は、変化への抵抗として以下のようにモデル化される：

$$\frac{d\theta_i}{dt} = (1 - b_{sq,i}) \cdot \omega_i \quad (4.1)$$

ここで  $\omega_i$  はバイアスなしの角速度である。

##### 確証バイアス (Confirmation Bias)

確証バイアス  $b_{cf,i} \in [-1, 1]$  は、自己判断への過信として協調係数を修正する：

$$k'_i = k_i \cdot (1 + b_{cf,i} \cdot 0.5) \quad (4.2)$$

##### 狭い視野 (Narrow Framing)

狭い視野  $b_{nf,i} \in [0, 1]$  は、全体目標への配慮の減少として表現される：

$$\mathbf{v}'_{tilda,i} = (1 - b_{nf,i}) \cdot \mathbf{v}_{tilda,i} \quad (4.3)$$

## 4.2 シミュレーション実験

### 4.2.1 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

### 4.2.2 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

## 4.3 結果と考察

### 4.3.1 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（図??を参照）。

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 [12]、政策変化を受け入れる「臨界点」として解釈できる。

### 4.3.2 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。

### 4.3.3 狭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。

## 4.4 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. 狭い視野は**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 生成 AI の位置づけ

#### 5.1.1 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**対処**する

#### 5.1.2 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。

### 5.2 認知バイアスへの対話的介入

#### 5.2.1 「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。

### 5.2.2 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

### 5.2.3 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。

## 5.3 ZK-SNARKs 概念の援用

### 5.3.1 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

### 5.3.2 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

## 5.4 LLM as a Judge による実装

### 5.4.1 Constitutional AI と市民討議

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



### 5.4.2 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

### 5.4.3 控訴プロセスによる人間介入

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

## 5.5 ZK-SNARKs 型政策評価のアーキテクチャ

### 5.5.1 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、秘匿化处理

**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

### 5.5.2 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。

## 5.6 小括

本章では、生成 AI と人間の協調的關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 狭い視野：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：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
- ・ 狭い視野への対処：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
- ・ 長期的影響：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、運用ルールの方策
3. 社会的受容：市民の理解と信頼の獲得

### 7.4.3 生成 AI と人間の協調的關係性の継続的検証

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 と人間の協調的關係性を探求した。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 Russo 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、 $\epsilon_1, \epsilon_2$  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	社会影响、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