都市・地域計画におけるAI技術導入に伴うアカウンタビリティの課題と要件

羽鳥　剛史

[1] 正会員　愛媛大学准教授　社会共創学部（〒790-8577 愛媛県松山市文京町3番）
E-mail: hatori@cee.ehime-u.ac.jp (Corresponding Author)

都市・地域計画の分野において，AIやビックデータ等の先端技術を活用した様々な取り組みが進められている．本研究では，都市・地域計画におけるAI技術導入に伴うアルゴリズムのアカウンタビリティ(algorithmic accountability)問題を取り上げ，アカウンタビリティの概念や課題を整理すると共に，アカウンタビリティを確保するための社会的・制度的要件について考察する．特に，アルゴリズムに関わるアカウンタビリティ概念の構造が，特定の手続き規則に則って個別の意思決定プロセスを正当化する個別的なアプローチと，社会の中で意思決定を正当化するための根拠や基準の正統性を確保するシステミック（全身的）アプローチにより成り立つことを指摘し，それぞれのシステム要件や機能を考察する．

Key Words : *algorithmic accountability, artificial intelligence, reflective responsibility, legitimacy*

1. はじめに

Society5.0（超スマート社会）やスーパーシティ構想において，人工知能（AI）やビックデータが社会変革を導く中心的な技術として位置付けられている．バルセロナ市をはじめ，スマートシティの先進都市では，AI技術を組み込んだ都市OS(operation system)を基盤として，都市・交通問題の解決を目指す取り組みが既に本格化しており，国内のいくつかの都市においても実験的な取り組みが始められている．

スマートシティに関わる政策提言の多くは，AIやビックデータの積極的な活用と同時に，人間中心的な価値に基づいて，市民の主体性や社会的な包摂性の理念を提唱している．しかし，これらの構想が描くように，市民が主体的な役割を発揮しつつ，AIやビックデータを活用し，都市・地域の課題解決を図っていくことは必ずしも容易ではない．特に，機械学習をはじめ，今日主流になりつつあるAIアルゴリズムは，大量のデータから意思決定規則（「もし…（条件）ならば…（結果）」という条件付言明）を自律的に生成するため，仮にその判断が偏見や不公正を含んでいたとしても，その理由や根拠を十分に説明できないという課題を抱えている．さらに，その意思決定システムには，アルゴリズム設計者，データ保持者，政策立案者，サービス提供者，サービス利用者をはじめ，多様な関係主体が関与しており，関係主体の間で意思決定の責任を帰属・分配することが難しいことも指摘されている．そのため，このようにアルゴリズムの判断過程が不透明な技術に依拠した社会では，社会的な意思決定への市民の主体的な参加が制限され，社会的な排除を助長する可能性が懸念されている[1),2)].

こうした中， AI技術の開発・利用に関わる政策議論において，アルゴリズムを活用した意思決定(algorithmic decision-making)のアカウンタビリティ（説明責任）をいかにして確保するかという問題が提起されている[3),4)]. 国内外の公的機関や学術組織等のAIガイドラインや倫理規定において，アカウンタビリティの要件は，AI技術の開発や利用において遵守すべき基本原則に位置付けられている．例えば，欧州委員会のAIガイドラインでは，人間中心的なAIの開発に関わる倫理原則として，慈善性（善をなすこと）と無危害性，自律性（人間の主体性を維持すること），正義（公平であること），説明可能性（透明性の中で操作すること）を挙げている．我が国の「人間中心のAI社会原則」においても，1) 人間中心の原則，2) 教育・リテラシーの原則，3) プライバシー確保の原則，4) セキュリティ確保の原則，5) 公正競争確保の原則，6) 公平性，説明責任及び透明性の原則，7) イノベーションの7つの原則が定められており，その中で「説明責任」に関して「AI を利用しているという事実，AI に利用されるデータの取得方法や使用方法，AI の動作結果の適切性を担保する仕組みなど，用途や状況に応じた

適切な説明が得られなければならない」と述べられている[5].

この様に，AIの開発・利用において遵守すべき原則として，アルゴリズムに関わるアカウンタビリティ(algorithmic accountability)の要件が定められているが，土木計画学の分野において，AI技術を活用する上でいかにしてアルゴリズムに関するアカウンタビリティを確保するかについては十分に検討されていない．そこで本研究では，都市・地域計画におけるAI技術導入に伴うアルゴリズムのアカウンタビリティ問題を取り上げ，アカウンタビリティの概念や課題を整理すると共に，アカウンタビリティを確保するための社会的・制度的要件について考察する．

2. アカウンタビリティの概念と課題

(1) アルゴリズムのアカウンタビリティ概念

アカウンタビリティ概念に関して，会計学，社会学，政治学，経済学，心理学等，多くの学問分野において研究が蓄積されている．その定義や概念は多義的であるが，伝統的には，委託者と受託者との間の 2 者関係を前提として発達してきた[6]．すなわち，Bovens 等[7]によれば，アカウンタビリティの委託ー受託関係は，「受託者Aが委託者 B に対して自己の行為 C を正当化する義務を果たし，もし委託者Bが受託者Aの正当化が不十分であることを発見した場合，一定の制裁を受ける可能性がある時，受託者 A は委託者 B に対して行為 C に関して説明可能である」という形式により定義される．Binns[8]は，このアカウンタビリティ概念をアルゴリズムによる意思決定の文脈において捉え直し，アルゴリズムに関わるアカウンタビリティにおける意思決定者とその対象者の関係について次のように記述している．すなわち，意思決定者は，アルゴリズムによる意思決定システムの設計と運用に関する理由及び説明を意思決定の対象者に提供しなければならない．意思決定の対象者は，この正当化が適切であるかどうかを判断することができ，適切でない場合には，意思決定者は何らかの制裁を受けたり，特定の意思決定の撤回や修正を余儀なくされる可能性がある．

現実のアルゴリズムによる意思決定は，特定の委託者と受託者の間で実施されることは稀であり，不特定多数の市民（委託者）から構成される社会の中で，複数の意思決定者（受託者）間の連携の中で生産される場合が一般的である．この点を踏まえて，アルゴリズムによる意思決定プロセスが社会の規範や手続きに則っているか否かという観点から，アルゴリズムに関わるアカウンタビリティを広義に捉える見方も提示されている．例えば，OECDのAI原則に拠れば，アカウンタビリティは「組織や個人が，自らの役割や適用される規制の枠組みに従って，設計，開発，運用，または導入した AI システムが，そのライフサイクルを通じて適切に機能することを確保し，自らの行動や意思決定プロセスを通じてこれを実証すること」を指す．また，Kroll[9]は「政策決定に至るプロセスが法的，政治的，社会的な規範と整合的であり，この事実が公に明白であるならば，そのプロセス及び意思決定主体は説明可能である」と指摘している．Koene 等 [3]は，アルゴリズムに関わるアカウンタビリティを「法的・倫理的な義務，方針，手順，仕組みにコミットし，社内外のステークホルダーに倫理的な実行を説明・実証し，適切な行動が取れなかった場合には是正することを含む，ガバナンス構造に集約される一連のメカニズム，実践，属性」と定義している．

(2) アルゴリズムのアカウンタビリティ課題

従来の研究より，意思決定にアルゴリズムを導入することにより，一般の人々に対するアカウンタビリティがより困難になる可能性が指摘されている[10),11)]．既往研究の議論より，アルゴリズムに関するアカウンタビリティの問題は，1) アルゴリズムの秘匿性，2) アルゴリズムの解釈可能性，3) 異質性とネットワーク性，4) 動的変容の 4 つに整理できる．第 1 に，アルゴリズムのソースコードは，国家機密や知的所有権の下，一般には秘匿されている場合が少なくない．その結果，一般市民に対してアルゴリズムの判断や根拠の妥当性を検証・説明することが制限される可能性がある．第 2 に，アルゴリズムの判断過程がブラックボックス化しているため，その判断に至った理由や根拠を解釈できない可能性がある．特に，設計者がアルゴリズムの意思決定規則を定める「ルール・ベース(rule-based)」による方法と異なり，機械学習のように，大量のデータから意思決定規則を生成する「ルール学習(rule-learning)」による方法では，設計者でさえもアルゴリズムがなぜその判断を下したのかを十分に説明できない問題が指摘されている[12)]．近年では，AI 判断の根拠を明確化する「説明能力のある AI(explainable AI)」の技術開発も進められているが，未だ発展段階にあり，AI 判断の根拠を一義的に理由付けすることは困難である．第 3 に，アルゴリズムは，通常，数多くのアルゴリズムから構成されるシステムに組み込まれており，システム全体として機能を発揮する．個々のアルゴリズムやシステム全体の設計，開発，照査，修正には，多くの関係者が関与する．さらに，アルゴリズムのシステムは，設計者等の設計・運用環境，利用者の使用環境，ソフト・ハードウェアの技術環境，法制度や社会規範を含む，異質な人的・物的ネットワークが錯綜する複雑な社会技術的な文脈に埋め込まれている．こうした状況においてシステムが下した判断の根拠やその責任を同定することは極めて難しい．第 4 に，アルゴリズムは，環境に

表-1 アルゴリズムのアカウンタビリティシステム

	システムⅠ	システムⅡ
アプローチ	個別的アプローチ	システミックアプローチ
システム機能	意思決定の正当化	意思決定基準の正統性の確保
システム要件	手続き的規則性	認識的・道徳的正統性
責任関係	責任の帰属	反省的責任（開放性，行為への親和性，目的論的志向性）

併せて動的に変容する．多くのアルゴリズムは，入力内容や出力パフォーマンスに応じて，そのコードを常に書き換えるようにプログラムされている．また，アルゴリズムの設計・修正過程には，ランダム性が組み込まれている場合が少なくなく，その過程を完全に予想・制御することは困難である．

これらのアカウンタビリティ課題に対して，意思決定プロセスの透明性(transparency)の向上を求める議論もある．既存の AI ガイドラインや倫理規定においても，透明性の原則が提示されている．しかし，意思決定プロセスの透明性だけでは，アカウンタビリティを確保することは難しい点が指摘されている[13),14)]．第 1 に，上述した通り，アルゴリズムのソースコードは，国家機密や知的所有権により秘匿される必要がある場合が少なくない．ソースコードを公開することにより，不特定多数の不当な介入に対する意思決定システムの脆弱性が高まる可能性もある．第 2 に，ソースコードを公開しただけでは，その機能や特性を検証する上では不十分である．特に，アルゴリズムに関する専門的な知識を持たない一般の市民は，ソースコードを閲覧できたとしても，その意味を理解することは容易ではない．第 3 に，アルゴリズムの判断過程にランダム性が組み込まれているため，仮に透明性を確保したとしても，それだけではその意思決定に対する恣意的な操作性を除去することは出来ない可能性がある．最後に，アルゴリズムは動的に変容するため，ある一時点のソースコードを公開しても，アルゴリズムの更新内容やその妥当性を評価することは難しい．これらの理由により，アルゴリズムによる意思決定プロセスの透明性を確保するだけでは，その妥当性を検証することは難しいと言える．

3. アカウンタビリティの構造と要件

(1) アカウンタビリティシステムの構造

アルゴリズムに関するアカウンタビリティの理念系を1つのシステムとして表現し，その構造と機能を分析する．アルゴリズムによる意思決定システムでは，複数主体間の複雑な委託ー受託関係が錯綜し，アルゴリズムを媒介した様々な意思決定が生産・再生産される．Kaminski[15)]の議論を踏まえると，こうしたアルゴリズムによる意思決定を規律付けるアプローチは，個別の意思決定の適正性を評価する個別的アプローチと，社会の中で意思決定システムの妥当性を評価するシステミック（全身的）アプローチに大別される．本研究では，この分類を参照し，**表-1**に示すように，アカウンタビリティシステムを2つのシステム（以下，アカウンタビリティシステムIとアカウンタビリティシステムIIと呼ぶ）から構成されるものと捉える．

第1に，アカウンタビリティシステムIでは，アルゴリズムを用いた意思決定が適正なプロセスに則っているかどうかが評価される．そこで，意思決定者は，公平性等の手続き基準に基づいて，アルゴリズムを介して入力情報から特定の判断を出力する論理関係を正当化(justify)し，自らの責任を果たすことが求められる．アカウンタビリティシステムIは，一定の評価基準を所与として，アルゴリズムを用いた意思決定を正当化し，その責任帰属を確立する役割を果たす．

第2に，アカウンタビリティシステムIIでは，アルゴリズムを用いた意思決定を正当化するための根拠や基準が社会的文脈の中で正統性(legitimacy)を持ち得るかどうかが評価される．そこでは，アルゴリズムを用いた意思決定に関わる関係主体間の責任関係のあり方が吟味される．Stahl[16)]は，この様に，関係主体の責任関係自体の妥当性を対象とした責任概念を反省的責任(reflective responsibility)と呼称し，その特徴として，1) 責任の帰属が広く社会全体に開かれているか否かという「開放性(openness)」，2) 責任の帰属により，人々の行動を改善できるか否かという「行為への親和性(affinity to action)」，3) 責任の帰属により，「望ましい社会」や「良い生活」等の社会的な目的を達成できるか否かという「目的論的志向性(teleological orientation)」という3つを挙げている．アカウンタビリティシステムIIは，アルゴリズムによる意思決定が依拠する認識論的・規範的基準を広い社会的文脈の中で検証することにより，こうした反省的責任を確立する役割を担う．

(2) アカウンタビリティシステムIの要件

都市・地域計画に関わる意思決定アルゴリズムは，当該エリアや個々のサービス利用者に関するデータを入力し，定められた計算手続きにより，都市問題の解析・予測や当該利用者の行動特性等に関わる一定の判断を出力する．アカウンタビリティシステムIでは，こうした意

思決定の要因やその相対的な重み付けを評価し，その意思決定プロセスが一定の手続き基準に従って進められているかどうかが問われる．こうしたアルゴリズム評価に関して，Doshi-Velez等[17]は，アルゴリズムを介して特定の入力が特定の出力にどのように影響を与えたかを定量化することにより，アルゴリズムの判断の妥当性について合理的な説明を行うことが出来ることを指摘している．

一方，アルゴリズムの手続き基準に関して，Kroll等[13]は，アルゴリズムを用いた意思決定プロセスが，そのアカウンタビリティを果たす上で準拠すべき条件として，「手続き的規則性(procedural regurality)」の基準を提示している．すなわち，手続き的規則性とは，「全ての人々に同様の手続きが適用され，かつ，その手続きが特定の人を不利にする方法で設計されていないことを関係者が理解できる」ことを要請しており，手続きの適用性と公平性の2つの基準から成り立つ．手続きの適用性に関して，Kroll等[13]は，暗号理論におけるゼロ知識証明の考え方に基づいて，意思決定者がインプットやアウトプット情報を公開することなく，事前にコミットしたアルゴリズムが実際に適用されたことを正当化・証明することを可能にする手法を提案している．一方，公平性の基準に関しては，機会学習やデータマイニング分野において，公平性配慮型データマイニング(fairness-aware data mining)等の手法が開発・検討されており，公平性，差別，中立性，独立性等の倫理的制約を考慮に入れた様々な分析方法が提案されている[18]．

アカウンタビリティシステムIでは，アルゴリズムを用いた意思決定プロセスがこれらの基準を満たしているかどうかを評価することにより，意思決定に関わる関係主体の責任を適切に帰属化させることが求められる．

(3) アカウンタビリティシステムIIの要件

アルゴリズムを用いた意思決定は，様々な認識論的・規範的な前提条件を内在化している[8]．第1に，アルゴリズムによる意思決定モデルは，モデルの一般化可能性，理論的なモデルとの整合性，因果関係と相関関係の相違等，認識論的な問題を孕んでいる．第2に，アルゴリズムによる意思決定モデルには，上述した通り，公平性等の倫理的制約が明示的あるいは暗黙の内に組み込まれている．さらに，都市・地域計画の分野においてアルゴリズムを用いた意思決定システムを導入する場合，自動運転車のトロッコ問題が典型的であるように，アルゴリズムの判断が一般の人々の道徳的判断に一定程度関与せざるを得ない．この意味において，アルゴリズムという技術は，それ自体において道徳的意義を持ち得ると考えらる．Binn[8]によれば，アルゴリズムに関わるアカウンタビリティでは，こうしたアルゴリズムを用いた意思決定モデルやプロセスが依拠している認識論的・道徳的な前提条件について，社会的な合意が形成されるかどうかについても主要な課題となる．アカウンタビリティシステムIIでは，アルゴリズムの認識論的基準や道徳的意義を社会全体の中で検証し，その正統性が評価される．また，意思決定システムにおける関係者間の責任関係が人々の行動を改善し，一定の社会的目的を達成しているかどうかを社会の中で吟味することにより，反省的責任の確立が目指される．その具体的な手法としては，アルゴリズムの設計者，運営者，政策立案者，一般の利用者を含む，多様な関係主体の意見をアルゴリズムの設計・運用にフィードバックさせる「構成的技術アセスメント(constructive technology assessement)」の方法が考えられる[19]．

以下では，アカウンタビリティシステムIIの機能的要件として，アルゴリズムが依拠する認識論的・道徳的基準の正統化の課題について述べる．

a) 認識的正統性の要件

古典的な認識論(epistemology)では，知識を「正当化された真なる信念(justified true belief)」と捉える伝統的な知識観がある[20),21]．すなわち，ある信念が知識であるのは，その信念が真であるだけでは十分ではなく，それが真理をもたらす十分な理由を持つという意味において，正当化されることが必要であると考えられてきた．この意味において，アルゴリズムの判断過程に関して，現実世界に関わる真理に接近する方法として正当化できるかどうかが問われる．特に，機会学習モデルが仮に現象の因果関係を説明できなくても，政策立案者はその予測能力が十分であればモデルの使用を許容する可能性もあり[22]，その利用にあたっては，こうした認識論的前提に関して社会の中で一定の合意を得ておく必要がある．この点に関連して，社会認識論(social epistemology)は，「多数の人々によって知識が探求される際の通常の状況において，知識の探求はどのように組織されるべきか」を問う[23]．社会認識論の観点から，社会全体において様々な関係者の認識論的正当化の根拠を相対的に吟味し，アルゴリズムの設計・運用に関わる認識論的基準の正統性を確保することが必要である．

b) 道徳的正統性の要件

都市・地域の問題解決にアルゴリズムを利用する場合，その判断が市民生活の様々な場面に浸透するため，そこでの道徳的判断に直接的・間接的な影響を及ぼすことは避けられない．この点に関連して，Verbeek[24]は，人間という主体と技術という客体を分離する近代主義的な主客二分法に疑義を呈し，ポスト現象学の立場から人間の道徳的判断が技術との相互作用の中で形成されることを強調する．そして，技術が道徳的判断に関与する媒介的役割に関して，予益原則（技術の使用者や使用によって影響を受ける人々に恩恵を与えるか），自律性の尊重（人々は技術の介入を自覚し，自律的な判断を保持でき

るか），無害性（人々のプライバシーは尊重されているか）等の道徳原則に照らして分析・評価する考え方を提案している．この点を踏まえると，都市・地域計画に関わる意思決定アルゴリズムが人々の経験や実践に対してどのような道徳的意義を有しているかを明らかにし，そうしたアルゴリズムの媒介的役割が自由や責任等の道徳的価値に照らして正統化できるかどうかが重要な課題となる．

4. まとめ

本研究では，都市・地域計画における AI 技術導入に伴うアルゴリズムのアカウンタビリティ問題を取り上げ，既往研究の知見に基づいて，アカウンタビリティの概念や課題を整理すると共に，アカウンタビリティを確保するための社会的・制度的要件について考察した．アカウンタビリティシステムの構造や機能については，1 つの仮説的モデルであり，その妥当性に関して，現実のアルゴリズムを巡る合意形成事例等を参照しつつ更なる検討を加える必要がある．

参考文献

1) Danaher, J.: The threat of algocracy: reality, resistance and accommodation, *Philosophy and Technology*, 29(3), pp.245-268, 2016.
2) Pasquale, F.: *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge, Massachusetts: Harvard University Press, 2015.
3) Koene, A., Clifton, C., Hatada, Y., Webb, H., & Richardson, R.: *A Governance Framework for Algorithmic Accountability and Transparency*, Brussels: European Parliamentary Research Service, 2019.
4) Diakopoulos, N.: Algorithmic accountability: journalistic investigation of computational power structures, *Digital Journalism*, 3(3), pp. 398-415, 2015.
5) 内閣府：人間中心の AI 社会原則，2018.
6) 越水一雄，羽鳥剛史，小林潔司：アカウンタビリティの構造と機能：研究展望，土木学会論文集 D, Vol.62, pp.304-323, 2006.
7) Bovens, M., Goodin, R. E., & Schillemans, T.: *The Oxford Handbook of Public Accountability*, Oxford: OUP Oxford, 2014.
8) Binns, R.: Algorithmic accountability and public reason, *Philosophy & Technology*, 31(4), pp.543–556, 2018.
9) Kroll, J.A.: *Accountable Algorithms*, PhD Thesis, Princeton University, 2015.
10) Nissenbaum, H.: How computer systems embody values, *Computer*, 34(3), pp.120–119, 2001.
11) Kitchin, R.: Thinking critically about and researching algorithms, *Information, Communication & Society*, 20(1), pp.14-29, 2017.
12) Stilgoe, J.E.Z.: Machine learning, social learning and the governance of self-driving cars, *Social Studies of Science*, 48 (1), pp. 25-56, 2018.
13) Kroll, J.A., Barocas, S., Felten, E.W., Reidenberg, J.R., Robinson, D. G., & Yu, H.: Accountable algorithms, *University of Pennsylvania Law Review*, 165, 633-705, 2016.
14) Ananny, M. & Crawford, K.: Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability, *New Media & Society*, 20, pp.973–989, 2016.
15) Kaminski, M.K.: Binary governance: lessons from the GDPR's approach to algorithmic accountability, *Southern California Law Review*, 92(6), 1529-1616, 2019.
16) Stahl, B.C.: Accountability and reflective responsibility in information systems, In Zielinski, C., Duquenoy, P., & Kimppa, K. (eds), *The Information Society: Emerging Landscapes*, pp. 51–68, New York: Springer, 2006.
17) Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., Scott, K., Schieber, S., Waldo, J., Weinberger, D., Weller, A., & Wood, A.: Accountability of AI under the law: the role of explanation, *Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center*, 2017.
18) Pedreschi, D., Ruggieri, S., & Turini, F.: Measuring Discrimination in Socially-Sensitive Decision Records, *Proceedings of the 2009 SIAM International Conference on Data Mining*, pp.581–92, 2009.
19) Rip, A., Misa, T., Schot, J. (eds.): *Managing Technology in Society: The Approach of Constructive Technology Assessment*, London: Pinter, 1995.
20) 戸田山和久：知識の哲学，産業図書，2002.
21) BonJour, L.: *The Structure of Empirical Knowledge*, Harvard University Press, 1985.
22) Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z.: Prediction policy problems, *The American Economic Review*, 105(5), pp.491–495, 2015.
23) Fuller, S.: *Social Epistemology*, Indiana University Press, 1988，小林傳司，調麻佐志，川崎勝，平川秀幸（訳）：科学が問われているーソーシャル・エピステモロジー，産業図書，2000.
24) Verbeek, P.P.: *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: University of Chicago Press, 2011，鈴木 俊洋 (訳): 技術の道徳化: 事物の道徳性を理解し設計する，法政大学出版局，2015.

(Received October 1, 2021)