社会科学論集　第 165・166 合併号　2022. 3

《翻　訳》

人工知能と透明性

── ブラックボックスをこじ開ける ──

トーマス・ヴィッシュマイヤー（著）
訳：栗島智明・小西葉子

目　次

〔要　旨〕

人工知能（以下，「AI」とする）の不透明性に関する主張は，ここ数年間で主要な政治的イシューへと発展した。一般に，利用者のプライバシー侵害を発見し，バイアスを見つけ出し，その他のありうる害悪を避けるためには，この〔AI という〕ブラックボックスをこじ開けることが必要不可欠だといわれる。しかし，はっきりとしていないのは，AI 透明性の要求がいかにして合理的な規制へと変換されうるか，という問題である。この章〔本稿は論文集のなかの一章として執筆された ── 訳者注〕では，AI 透明性をデザインすることは一般に想定されているほど困難ではないことを論じる。私たちの法システムはすでに，人間の決定という，部分的に不透明な意思決定システムをいかに解明するかという問題について，非常に多くの経験を積んできた。このことは，AI を規制しようとする人々にとって有利にはたらくだろう。このような経験を活かせば，AI 透明性に関する将来の立法およびそのために用いられる法的手段がどのように機能するかについて，法律家は現実的な見通しを立てることが可能である。

1　はじめに

【1】AI が社会に及ぼす影響力が増大するにつれて，こんにち，AI を搭載したシステム（AI-based systems. 以下，「AI 搭載システム」とする）を規制し，利用し，あるいはそれによって影響を受けている人々は，その技術について適切な理解を有すべきだという広範な合意が形成されてきている。絶え間なく出される一連の政策ペーパーや，国家計画戦略，専門家の提案，ステークホルダーの提言は，AI 透明性という観点から，この点をまとめている(1)。たしかに，AI は人間よりもはるかに素早くかつ正確にパターンを分析することができ，それゆえ，AI 搭載システムは，多くの場合，人間の知能で理解するにはあまりに複雑すぎる問題を解明するために用いられている(2)。この事実は忘れられてはならないが，透明性を促進するための技術が，AI 透明性に関する議論の中心に位置するわけではない。むしろ，AI の不透明性（opacity）の主張それ自体が，ここ数年のあいだの主要な政治的イシューになっている。それゆえ一般に，利用者のプライバシー（参照，マルシュ・本書所収「AI とデータ保護基本権」5 段落以下）への侵入を突き止め，バイアス／AI 差別（参照，ティシュビレック・同「AI と差別」19 段落）を発見し，またその他のありうる害悪（参照，ヘルムシュトリューヴァー・同「AI と不確実性のもとでの行政決定」45 段落，ラーデマッハー・同「AI と法執行」31 段落）を避けるために，この〔AI という〕ブラックボックスをこじ開けることが必要不可欠だと論じられている(3)。

【2】透明性は常に，データ保護の一般原則であり続けてきたが（EU 一般データ保護規則（以下，「GDPR」とする）5 条 1 項(a)），いまや世界中の立法者が，AI 搭載システムを含む，自動化された意思決定システム（automated decision-making systems（以下，「ADM」ないし「ADMs」とする））に対して特別な透明性の要求を課すという実験を始めている。2017 年，ニューヨーク市議会において「特定の人々へのサービスのターゲティング，制裁の賦課，またはポリシングの目的のため，アルゴリズムその他の自動化されたデータ処理システムの手段を用いている」すべての市の機関について，とりわけ「当該機関のウェブサイトにおいて該当するシステムのソースコードの公開」を義務付けることが提案された(4)。また 2018 年には，ドイツ情報自由監察官会議によって，ADM を利用している公的機関および民間アクターに対し，当該システムの「論理（logic）」や，インプットデータに対して適用される分類子（classifiers）および重み（weight），システム運用者の専門知識の程度について，詳細な情報公開を義務付ける法律を制定することが主張された(5)。さらに同じく 2018 年，ドイツの州司法大臣会議により設置されたタスクフォースにより，ADM

（1）　ここではいくつかを挙げるにとどめる。参照，National Science and Technology Council Committee on Technology (2016), OECD Global Science Forum (2016), Asilomar Conference (2017), European Parliament (2017), Harhoff et al. (2018), Singapore Personal Data Protection Commission (2018), Agency for Digital Italy (2018), House of Lords Select Committee on Artificial Intelligence (2018), Villani (2018), European Commission (2018) and Datenethikkommission (2018).

（2）　Bundesanstalt für Finanzdienstleistungsaufsicht (2018), pp. 144–145. さらに参照，Hermstrüwer〔本書所収「AI と不確実性のもとでの行政決定」〕, para 3, Hennemann〔本書所収「AI と競争法」〕, para 37.

（3）　この議論において重要な文献として，Mayer-Schönberger and Cukier (2013), pp. 176 et seq.; Zarsky (2013); Pasquale (2015); Burrell (2016); Diakopoulos (2016); Zweig (2016); Ananny and Crawford (2018).

（4）　当初の動議（Int. 1696–2017）では，上述の引用文が，ニューヨーク市行政法典 23–502 節に追加されるはずであった。しかし，最終的に可決された法律では，結局，市の機関が現在どのようにアルゴリズムを活用しているかについて研究するタスクフォースが設置されるにとどまった。立法プロセスの詳細な説明として，see legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0.

（5）　Konferenz der Informationsfreiheitsbeauftragten (2018), p. 4.

の使用について公衆に通知する一般的義務を導入することが提案された[6]。EU ではコミッションが 2018 年に「オンライン仲介サービスのビジネスユーザーのための公正性・透明性を向上するための規則」を導入し，そこでは順位付けアルゴリズム（ranking algorithms）についての開示義務が予定されている[7]。ドイツでも，「メディア仲介者（media intermediaries）」について同様の提案が議論されている（参照，クレンケ・本書所収「AI とソーシャルメディア」55 段落）[8]。ADM を利用する特定の金融活動については，通知義務がすでに導入されている（参照，シェンメル・同「AI と金融市場」25 段落）[9]。

【3】透明性を求める声は，一定程度では，新しいテクノロジーをめぐる議論に特有のものだといえるし，それに対する批判についても同様である[10]。ここでなされる批判は，第一に，技術的な実現可能性を疑問視する。一部の論者によると，透明性について意義ある規制を行おうとすれば，必ず失敗するのだという。その理由は，ジェナ・バレル（Jenna Burrell）の表現を借りれば，AI を用いたシステムの意思決定プロセスは，その固有の性質ゆえに「ある入力データが分類され，特定のアウトプットを生み出す方法や理由につき，その決定によって影響を受ける人間がほぼまったく理解しえないという意味で，不透明（opaque）」だからである（本章【10】以下を参照）[11]。第二の批判の潮流は，透明性に関する規制が前提としている規範的想定が誤っていると主張し，そこでの想定は，アカウンタビリティについての空想を生み出すだけであるという（本章【23】以下を参照）[12]。第三の批判は，本来的な問題は AI が透明性に欠いていることではないと主張する。この批判者たちによれば，問題はむしろ ── この議論はやや不明瞭なところがあるのだが ──，AI の「説明可能性（explainability）」，「明瞭性（intelligibility）」，「わかりやすさ（comprehensibility）」，「理解可能性（understandability）」ないし「予見可能性（foreseeability）」，逆に言えば，AI を用いたシステムの「秘密性（secrecy）」，「不可解性（inscrutability）」ないし「非・直観性（non-intuitiveness）」なのだという（本章【25】以下を参照）[13]。

【4】たしかに，AI を用いたシステムのソースコードやデータベースが閲覧可能だとしても，それは素人にとってそれほど意義あるものではないだろうし，また，情報公開法（sunshine laws）が存在するだけでアカウンタビリティが担保されるという保証はないだろう。しかし，そうはいっても，AI の透明性を追求する試みは必要であり，意義あるものである。第 29 条作業部会〔訳注：1995 年の EU 個人データ保護指令のもと，EU における個人データ保護の指針を示してきた作業部会を指す〕のステートメントをパラフレーズするならば，技術的複雑性は，重要な情報を提供しないことの言い訳にはならない[14]。〔AI の〕透明性は，その価値において第一義的に手段に過ぎないかもしれないが，そうだとしても，あらゆるアカ

（6） Arbeitsgruppe “Digitaler Neustart” (2018), p. 7.
（7） COM/2018/238 final — 2018/0112 (COD).
（8） Rundfunkkommission der Länder (2018), pp. 25–26.
（9） 2018 年 1 月時点で，ドイツにおいて投資サービスを提供する事業者は，ドイツ証券取引法（WpHG）80 条 2 項にいうアルゴリズム取引を用いている場合には，通知義務を負う。これは，EU の第 2 次金融商品市場指令（MiFID II）の国内実施規定である。
（10） 参照，Mittelstadt et al. (2016), p. 6：「ナイーブなことに，透明性はしばしば，新たな技術から生じる倫理的問題に対する万能薬のように扱われている」。同旨のものとして，Neyland (2016), pp. 50 et seq.; Crawford (2016), pp. 77 et seq.
（11） Burrell (2016), p. 1.
（12） 参照，Ananny and Crawford (2018), p. 983. この点に関する広範囲にわたる議論について参照，Tsoukas (1997); Heald (2006), pp. 25–43; Costas and Grey (2016), p. 52; Fenster (2017).
（13） 関連文献を含め参照，Selbst and Barocas (2018), pp. 1089–1090.
（14） Article 29 Data Protection Working Party (2018), p. 14.

ウンタビリティの枠組みにおいて欠くことのできない要素なのである。とりわけ公的機関における意思決定過程は，ドイツ連邦憲法裁判所の表現を借りれば，「一般的に可視的であり，かつ，理解可能で」なければならない(15)。この領域において透明性の要求は，究極的には法の支配および民主的統治の原理によって基礎づけられる。しかしまた，透明性は民間セクターにおいても極めて重要な機能を果たす。民間セクターでは，透明性について規制を行うことは，市場への参入コストを低下させ，また，事業者・消費者間などにおける構造的かつ有害な情報の非対称性を再度調整することへとつながる。それゆえ，AI を用いたテクノロジーが用いられる市場がますます増えているこんにち，民間の AI について透明性を要求することは，そこでの競争を作り出し，あるいは維持することへとつながるのである(16)。

【5】しかし，AI 透明性の要求は，いかにして合理的な規制へと変換しうるだろうか？　私見によれば，AI 透明性をデザインすることは複雑かもしれないが，一般に想定されているほど困難ではない。私たちの法システムはすでに，人間の決定（human decisions）という，部分的に不透明な意思決定システムをいかに解明するかという問題について，非常に多くの経験を積んできた。このことは，AI を規制しようとする者にとって有利にはたらくだろう。この経験を活かせば，AI 透明性に関する将来の立法およびそのために用いられる法的手段がどのように機能するかについて，法律家は現実的な見通しを立てることが可能である。

【6】機能（*functions*）という点では，人間〔の頭脳〕というブラックボックスをこじ開けようとする過去の取り組みによって，「完全な」透明性は可能でもなければ望ましくもない，ということが分かっている。むしろ，透明性に関する規制は，広い意味において，次の 2 つのことを目標とすべきである。すなわち，第一に，市民・消費者・メディア・議会・規制機関・裁判所といった，様々な利害関係者の集団に対して知識（knowledge）を提供するためには，それぞれの集団において，決定に関して行政的ないし司法的な審査を行うか，その他の方法により意思決定者に説明責任を課す必要がある。第二に，透明性に関する規制は，決定により直接的に影響を受ける者に対し，行為主体性（エージェンシー）の感覚を与えるものでなければならない。そのことによって，意思決定プロセスに対する信頼が生まれるのである。

【7】規制の道具立て（*regulatory toolbox*）という点では，透明性に関する規制は，システムおよびその運用についての情報を提供することもありうるし，あるいは，みずからの決定について関係者に説明することをシステムの運用者に要求することもありうる。このいずれの規制のタイプであっても，上述の機能を果たすことにつながるのであり，ゆえにそれは AI アカウンタビリティを高めることになる。しかし，この義務に応じる際，または，当該義務に対応する権利行使がなされる際に，システムの運用者が提供すべき情報の質や種類は，両者で異なってくる。

【8】〔上述の〕機能および道具立てについて現実的な態度を採るならば，AI 透明性に関する規制においてしばしば登場する 2 つの誤った考えを避けることができるだろう。以下 **2** において示されるように，AI に関する知識の創造を規制困難な事例としている要因は，第一義的には，認識上の

(15)　Bundesverfassungsgericht 2 BvR 2134, 2159/92 'Maastricht' (12 October 1993), BVerfGE 89, p. 185; 2 BvR 1877/97 and 50/98 'Euro' (31 March 1998), BVerfGE 97, p. 369. 透明性の価値につき，Scherzberg (2000), pp. 291 et seq., 320 et seq., 336 et seq.; Gusy (2012), § 23 paras 18 et seq.; Scherzberg (2013), § 49 paras 13 et seq.

(16)　Cf. CJEU C-92/11 'RWE Vertrieb AG v Verbraucherzentrale Nordrhein-Westfalen eV' (21 March 2013), ECLI: EU: C: 2013: 180; CJEU C-26/13 'Árpád Kásler and Hajnalka Káslerné Rábai v OTP Jelzálogbank Zrt' (30 April 2014), ECLI: EU: C: 2014: 282. See also Busch (2016). 米国の私法およびプライバシー法における開示義務に対して批判的な視点を提示するものとして，Ben-Shahar and Schneider (2011) and Ben-Shahar and Chilton (2016).

理由（epistemic reasons）――つまりブラックボックス問題 ―― ではない。そうではなくて，システム運用者は AI の秘密性について正当な利益を有しており，この領域における個人の情報アクセス権がデリケートな衡量の問題となってしまうのは，とりわけこの点に由来しているのである（しかし，説明する義務に関しては，それと同程度にデリケートな問題とはいえない）。たしかに，AI 搭載システムのなかには，自分自身を「説明する」ことが不可能なものも存在する。つまりそこでは，下される決定とその要因との因果関係をうまく説明することができない。しかしその事実は，個人の行為主体性（エージェンシー）〔の感覚〕という観点からすれば，それほど重要ではない。それよりもむしろ問題なのは，システムおよびその運用者がいまでも提供しうる ―― あるいは実際に提供している ―― 情報について，現時点では，規制機関および裁判所がそれを適切に処理するためのリソースを欠いている，という事実である。それゆえ，例えば最近の説明可能な AI（Explainable AI（XAI），以下「説明可能 AI」とする）に関するイニシアティブにおいても，AI の「説明可能性」は，テクノロジーのデザインに関する問題であるとされていたが，実際はそうではない。以下 **3** において議論する通り，AI の「説明可能性」とは，第一に，そして何よりも，行政および司法に対する制度的な挑戦なのである。このことを前提に，以下 **4** では，AI の透明性に関する意義ある規制をするための今後の道筋を素描する。

2　情報アクセスによる知識の創出

【9】現代の社会組織における認識レベルでの制約のもとで，知識（knowledge）は，孤独のうちに生み出されるのではなく，むしろ，批判的かつ開かれた議論のなかで生み出される[17]。社会の内部において，開放性（openness）は創造性を可能とし，批判を促進し，また，認知上その他のバイアスを防ぐのに役立つ。また政府の領域では，開放性によって応答性（responsivity）や効率，アカウンタビリティが担保されるのであり，ゆえに開放性は良い統治に資する要因といえる[18]。開放性を生み出し，もって知識の増大を促進させるための標準的な規制ツールとして，個人の情報へのアクセス権，開示（disclosure）を義務付けるルール，あるいは，公的機関に調査権限を認めることなどが考えられる[19]。AI の分野においても，そのような権利および義務 ―― これらは本稿では「情報へのアクセス規制」という大きな概念のなかにグループ化される ―― を認めることで透明性が促進されうるかは，何よりもまず，AI テクノロジーが不透明であるという主張のいかんにかかっている（**2.1**）。上述のような様々なルールが必要的かつ比例的な方法でデザインされうるか，という問題は，透明性に関する規制の存在により不利益を被る者が有している規範的な地位を考慮して考えられなければならない（**2.2**）。そして最後に，情報へのアクセス規制についての法令およびその及ぶ範囲について考える場合には，プラグマティックな考慮をすべきである（**2.3**）。そこではとりわけ，当該規制によって得られる知識を処理するために，個別の利害関係者がどの程度の能力を有しているかについて考える必要がある。

2.1　認識レベルでの制約

【10】AI 透明性に関する今日の議論では，認識レベルでの制約が強調されることが多い。たしかに，AI に基づいた意思決定システムは，外部の観察者に対して認識上の制約があることを意識させるし，実際，素人の視点からすれば，そのような制約が存在することは明らかである。コンピューターのコードを理解できるのはわずかな人々に限

(17)　Merton (1968), pp. 71–72.
(18)　参照，Fassbender (2006), § 76 para 2; Hood and Heald (2006); Florini (2007).
(19)　時として，透明性に関する規制は情報アクセス権を個人に認めることと同義と考えられている。しかし実際には，そのような個人の権利は，社会における情報フローを支配する規制構造の幅広い枠組みの中の一要素に過ぎない。これについては参照，本書第 4 章（Part 4）。

られるし，大多数の人々は，極めて初歩的な機械学習アルゴリズムでさえ分析することができない。同じことが，公務員についてもあてはまる。さらに，IT 業界における専門分化の程度が高まっているため，IT 分野において訓練を受けた専門家ですら，AI に基づく複雑なシステムで用いられている手法を完全には理解できないことがある(20)。

【11】たしかに，〔AI を認識することには〕困難が伴う。しかし，AI の背後の技術や論理が，実際には全体として極めてよく理解されているという事実が見逃されてはならない。まず，AI の定義が問題となるが，ここでは，様々な統計的手法およびアルゴリズムを適用する行為，しかもその際，潜在的に極めて多く存在し，かつ，内実において多種多様なデータ（いわゆる「ビッグデータ」）の習得を伴い，ゆえに一般的に機械学習という名で示されるものとして，定義しておこう。実際，このように定義された AI の背後にある技術や論理は，例えば環境変動や，ナノテクノロジー，金融市場といった自然現象・社会現象と比べても，はるかによく理解されているのである。それだけでなく，過去の数十年間のうちに，ソフトウェア産業において，複雑なシステムをテストするために通常用いる様々な統制ツールが開発されてきた(21)（ここでいう複雑なシステムには，AI 搭載システムも含まれる(22)）。このテストのために，検査官たる専門家チームは，コードをチェックしたり，基礎にあるアルゴリズムを再現（reconstruct）したり，訓練プロセスを分析したり，訓練・結果のデータベースを評価したり，あるいはダミーデータを流し込むこむことも可能である(23)。

【12】このように，AI についての私たちの理解は理論上，非常に進んできている。にもかかわらず，AI 透明性に関する研究においては，AI 検査（AI audits）が現実において直面している三つの深刻な問題点が指摘される。第一の問題点は，データ処理の量・多様性・速度が上がれば上がるほど，システムの作動を理解し，予測したり，コンピューター処理の相関関係を再現することがますます困難となる，という事態である(24)。しかし，この問題は AI 搭載システムに特有のものではない。こんにちでは，完全に決定的な（deterministic）プログラムであっても，潜在的には無限にある変数と属性を，極めて短い時間で処理することができる。さらに，多くの事例において，コンピューター技術それ自体が，この問題への解決策となる。というのも，少なくとも部分的には，〔プログラムの〕検査それ自体が自動化されうるためである。検査のために特化された AI が用いられることもあるだろう(25)。それゆえ，データを基礎とするシステムが人間にとって容易にアクセス可能か否かは，データの「大きさ（bigness）」によるものではない。むしろ，データがシステムのなかでどのように処理されているか，つまり，アルゴリズム

（20）　AI 搭載システムの「境界線」を確定することが困難であり，それが透明性の問いをより複雑なものとしている点につき参照，Kaye (2018), para 3; Ananny and Crawford (2018), p. 983.

（21）　AI に特有の文脈を超えて，エラーの発見と回避はこんにち，洗練された研究活動および広範な技術標準化の試みの対象となっている。参照，国際標準 ISO/IEC 25000 'Software engineering — Software product Quality Requirements and Evaluation (SQuaRE)'. これは，ISO/IEC JTC 1/SC 07 Software and systems engineering によって制定されたものである。なお，そのドイツ版である 'DIN ISO/IEC 25000 Software-Engineering — Quality Criteria and Evaluation of Software Products (SQuaRE) — Guideline for SQuaRE' は，ドイツ規格協会（DIN）の NA 043 Information Technology and Applications Standards Committee (NIA) によって制定されている。

（22）　この文脈で頻繁に参照される論文として，Sandvig et al. (2014), pp. 1 et seq.

（23）　最近になって，ミュンヘン再保険（Munich Re）とドイツ人工知能センター（DFKI）が共同で，オンライン上の不正支払いの探知目的で AI を使用するスタートアップ企業が使用する技術の検査を行った。この検査では，用いられるデータや，基礎にあるアルゴリズム，統計モデル，企業の IT インフラのチェックがそれぞれ行われたが，それは無事に完了し，当該スタートアップの保険は結果的に認められた。

（24）　Tutt (2017), pp. 89–90 は，IBM およびテスラの専門家が，企業のシステムの誤作動について事後的に理由を再現しえなかったいくつかの事例を描いている。

（25）　参照，前掲注（2）。

の処理方法こそが重要なのである。

【13】第二の問題の核にあるのは，アルゴリズムそれ自体である。AI 搭載システムで用いられるアルゴリズムのうち少なくとも一部についていえば，特定のインプットを事後的に特定のアウトプットと結び付けたり，その逆をすることは事実上，不可能である（理論的に不可能とまではいえないにせよ）[26]。しばしば，そのような因果関係による説明は，機械学習アルゴリズムを用いている AI 搭載システムにおいて不可能であると考えられている。しかし，機械学習は単一の概念ではなく，様々な技術から成り立っていることに注意しなければならない。すなわちそこでは，伝統的な線形回帰モデルが用いられることもあれば，サポートベクターマシン（SVM）や，決定木学習アルゴリズムが，様々なニューラルネットワークについて用いられることもある。特定のインプットとアウトプットのあいだに事後的な因果関係を見出すことの困難さは，これらの技術ごとにまったく異なる[27]。例えば，決定木学習アルゴリズムは，財務予測や融資申込プログラムなど，回帰問題と分類問題のいずれにも用いることが可能であるが，そこでは〔決定〕木が一度構築されれば，因果関係を説明することが可能となる。しかし他方，主として言語・画像認識や自然言語処理などのパターン認識のために用いられる人工ニューラルネットワークについては，必ずしもそれは当てはまらない。ここにおいて，再現の問題は深刻なものとなる。というのも，仮にシステムの作動に関する情報が完全に与えられたとしても，特定の決定を事後的に分析することで，人間にとって容易に理解できるような直線的な因果関係を導き出すことは不可能だからである。代わりに，事後的な分析によって得られるのは，多数の加重変数に関する確率関数のかたちをとった決定ルールである。この決定ルールはさらに，計算時間を最適化するために修正されることもある。仮にこの関数を書き出したり，再度計算することが理論的には可能だとしても，「機械学習の高次元的性格における数学的最適化と，人間の感覚に即した推論の要求および意味解釈スタイルとのあいだのミスマッチ」[28] が残ることになる。

【14】ニューラルネットワークの「不透明性」に関する主張は，エンジニアの観点からすると不満に思われるかもしれない。しかし，ニューラルネットワークで因果関係を特定することが困難だからといって，ニューラルネットワークについての（強制／任意の）AI 検査が無用であるということにはならない。まして，ニューラルネットワーク以外の AI についていえば，検査が不要ということはまったくない。なぜなら，システムおよびその作動方法についての情報を集める可能性がいまだ残されているからである。データを収集し，様々な仮説を確かめることによって，不透明な現象についての説明をより多く探し出すというのは，知識を生み出し，科学や社会の領域のデータを理解する際の，ごく標準的な方法である。そして，後述するように，これは情報アクセス規制の主要な目的のうちの一つである（後述 **2.3**）。

【15】第三の困難は，一般的に AI の進化が非常にダイナミックである点，とりわけ，AI 搭載システムのパフォーマンスがダイナミックに伸びてきている点に関わる[29]。IBM のワトソンを模範として，フィードバック・ループをビルトインした，AI 搭載システムの数はますます増えている。このフィードバック・ループにより，ユーザーへのアルゴリズムの影響に呼応するかたちで，変数の重みづけを常に調整しつづけることが可能と

（26） 以下の記述につき，さしあたり Hildebrandt（2011），pp. 375 et seq.; van Otterlo（2013），pp. 41 et seq.; Leese（2014），pp. 494 et seq.; Burrell（2016），pp. 1 et seq.; Tutt（2017），pp. 83 et seq.

（27） 議論の詳細につき Bundesanstalt für Finanzdienstleistungsaufsicht（2018），pp. 188 et seq.

（28） Burrell（2016），p. 2.

（29） 従来型のアルゴリズムであっても，その多くでは常にアップデートが繰り返されており，ゆえに事後的な評価は困難である。参照，Schwartz（2015）.

なる(30)。それゆえ観念上は，システムが作動することによって常に，システムそれ自体が訓練され，アップデートされることになる。ゆえに，t_0 において x のインパクトをもたらした要素が，t_1 においては y という異なった決定に至ることがありうる。このようなダイナミックなアーキテクチャにおいて，〔因果関係の〕説明はほんの一時的にしか妥当しない。それゆえ，「仮にアルゴリズムのソースコードや，その訓練データセット，さらに試験データのすべてが可視化されたとしても，当該アルゴリズムの機能についての特定のスナップショットが得られるに過ぎない」(31)。しかし，このことによって，AI 検査の可能性が否定されるわけではないし，情報アクセス規制が必要なくなるわけでもない。というのも，一瞬のスナップショットからでも，貴重な情報が得られるかもしれないからである。それゆえ，規制の観点からすれば，第三の困難は単に次のことを意味するに過ぎない。すなわち，透明性規制は，記録義務（documentary obligations）によって補完されなければならないが，そうすれば，情報アクセスの可能性は担保されうる，ということである（本章【48】を参照）。

2.2　規範的な制約

【16】以上述べたことからわかる通り，AI 搭載システムの運用者に対し，監督官庁や検査官を含む第三者を相手として，コードや，データベース，統計モデル，IT インフラなどにアクセスできるように義務付けることは不合理である，とアプリオリに断定するための強力な認識上の根拠は存在しない。しかし，現在のところ，そのように広範囲にわたる開示を求める情報アクセス規制は存在していない。AI 搭載システムが意思決定の目的で用いられている場合でも，ほとんどの法制度において，それが公に周知されるべきことが規定されることはないし，ましてや，コードにアクセスする権利など認められていない。ただし，よく議論がなされている例外として，フランスにおける，デジタル共和国のための 2016 年 10 月 7 日法が挙げられる。それによれば，「アルゴリズムに基づいて」国家機関が意思決定を行う場合，個人には，意思決定システムの「主要な特徴」についての情報を受け取る権利が認められる(32)。このフランスの法律は，1995 年 EU データ保護指令の 12a 条および 15 条の ADMs に関する規定から影響を受けており，当該条文は現在，EU 一般データ保護規則（GDPR）13〜15 条および 22 条によって代替されている。すなわち，GDPR 13 条 2 項(f)および同 14 条 2 項(g)によれば，EU 域内のあらゆるデータ運用者は，データ主体に対して「プロファイリングを含む自動処理決定の存在」および「使用された論理（logic）に関する意味のある情報」を提供しなければならない。またこれに対応して，GDPR 15 条 1 項(h)は，この情報にアクセスする権利をデータ主体に付与している。もっとも，GDPR に定められた権利は，同 22 条の意味における自動処理決定の場合に制限されている。すなわち，それは自動処理「のみ」に基づく決定であって，かつ，「法的効果」ないし「それと同様の重

(30)　参照，IBM (2018).

(31)　Ananny and Crawford (2018), p. 982. さらに参照，Diakopoulos (2016), p. 59.

(32)　Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique. そして Décret n° 2017-330 du 14 mars 2017 relatif aux droits des personnes faisant l'objet de décisions individuelles prises sur le fondement d'un traitement algorithmique（available at www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000034194929&categorieLien=cid）の 1 条に基づき創設された R311-3-1-2 条においてさらに詳しく述べられている通り，行政は，(1)アルゴリズム処理が意思決定にどのように貢献しており，また，その貢献はどの程度であるか，(2)いかなるデータがそこで処理され，そのデータはどこに由来するか，(3)いかなる変数を用いてデータが処理され，適切である場合には，その重みづけがどのようになされるか，(4)システムによってどのような操作（operations）がなされているか，に関する情報を提供しなければならない。これらの情報はすべて「理解できる形で提示されねばならず，かつ，法により保護された秘密を侵してはならない」。この法律には，国家安全保障上の例外もある。さらなる詳細につき Edwards and Veale (2017).

大な影響」を個人に対して有するものでなければ，適用されないのだ[33]。したがって，意思決定にあたって人間をサポートするだけの AI 搭載システムは，当該規定の範囲に含まれない[34]。さらに，GDPR 22 条が適用される場面においても，同 13 条〜15 条に基づいてデータ主体に対して提供されなければならない情報は制限されている。というのも，EU データ保護指令およびそれを実施する加盟国の国内法 ── そこには，2001 年のドイツ連邦データ保護法 6a 条および 28b 条が含まれる。なおこれらは，ほぼ同じ文言の 2018 年ドイツ連邦データ保護法 31 条および 38 条に取って代えられている ── についてであるが，個別事例の決定に関する詳細な情報は不必要であり，一般的には，システムの大まかなデザインの記述で十分である，と裁判所が判断しているためである[35]。データ運用者には，自動処理システムの一次データやコードなどへのアクセスを認める義務があると裁判所が判断した事例は存在しない。

【17】このような裁判所の判断が GDPR のもとで大きく変化するとは，あまり考えられない。その理由は，立法者や裁判所が，自動処理決定における透明性の重要性を低く見積もっているからではない。そうではなく，さらに進んで開示要求することを困難にさせる，重大な法的理由がいくつか存在するためである。そこに含まれるのは，知的財産権の保護であり，企業秘密・営業秘密（trade and business secrets）の保護である。これらは，GDPR の前文（recital）63 段 5 文において，透明性の限界として明示的に認められている[36]。さらに，プライバシーや法執行上の利益が，透明性の要求とぶつかり合うことがある[37]。「AI の秘密性」に関するこれらの主張は，これまでのところ体系的に扱われてきていない。この問題を完全に明らかにするためには，AI の公的な運用者と民間の運用者を区別し[38]，さらに，様々な種類の「法的な秘密（legal secrets）」（キム・シェッペル〔Kim Scheppele〕）を区別すること

（33） 詳細につき参照，von Lewinski (2018), paras 7 et seq.; Martini (2018), paras 16 et seq., 25 et seq.

（34） 参照，Martini and Nink (2017), pp. 3 and 7–8; Wachter et al. (2017), pp. 88, 92; Buchner (2018), para 16; Martini (2018), paras 16 et seq.; von Lewinski (2018), paras 16 et seq., 23 et seq., 26 et seq.

（35） データ保護指令の 12a 条につき参照，CJEU, C-141/12 and C-372/12, paras 50 et seq. クレジット・スコアリングについての特別な規定を含んでいた（現在も含んでいる）ドイツのデータ保護法に関して，連邦通常裁判所は，クレジット・スコアリング・システムのデザインに関する簡単な説明で十分であり，システム運用者は，高いクレジット・スコアとクレジットを受けられる蓋然性のあいだの大まかな関係性を説明するだけでよい，と判断した。個別事案における決定の詳細な情報やシステムに関する詳しい情報は必要でないとされたのである。Bundesgerichtshof VI ZR 156/13 (28. 1. 2014), BGHZ 200, 38, paras 25 et seq. 決定に用いられている論理についての適切な情報提供を行うために，具体的にいかなる情報が開示されなければならないかについて，学説上，実に様々な見解が存在する。異なる立場の概観として参照，Wischmeyer (2018a), pp. 50 et seq.

（36） 参照，Hoffmann-Riem (2017), pp. 32–33.「戦略的な対抗措置のリスク」につき参照，Hermstrüwer〔本書所収「AI と不確実性のもとでの行政決定」〕, paras 65–69.

（37） ここでは次の文献を挙げるにとどめる。Leese (2014), pp. 495 et seq.; Mittelstadt et al. (2016), p. 6.

（38） 法の支配および民主的統治の原理は，公的主体の透明性を義務づけ，行政の秘密性を制限する（本章【4】を参照）。これに対して，民間事業者に透明性を求めることは，事業者の基本権という観点からして正当化を必要とする。しかし，ここで問題となっている公共の利益の重大性および新しい技術がはらむリスクにかんがみれば，民間のシステム運用者の利益が完全に優位するということはほとんど考えられない。さらに立法者は，AI 搭載システムによって不利益を被る者の基本権を保護しなければならない。これは通常，新たな技術についての実効的な統制を可能とするような法律を作ることを意味する。しかし，この点で立法者が広範な裁量を有していることもまた認めなければならない。ただし，私有化された公共のスペース（いわゆる「パブリック・フォーラム」）を管理する企業や，何らかの理由で国家に近い権力を保持するに至った企業については，データ主体が有する基本権の水平的効力〔訳注：基本権の私人間効力を指す〕によって，通常よりも厳しい透明性規制が求められることがある。

が必要になるだろうが，それは，この章のなかでなしうることではない(39)。以下では，戦略上および信託上の秘密（strategic and fiduciary secrets）を保護する必要性が，AI 透明性の規制にどのような影響を与えているか，一般論を指摘するにとどめる。

【18】戦略上の秘密性は，競争における優位を得るために，情報の非対称性を固定させようとする。民間セクターにおいては，戦略上の秘密を守ることは例外ではなく，むしろ原則である。というのも，第三者に情報へのアクセスを認める義務を課すことは基本権の制約にあたり，正当化を必要とするからである。それゆえ，IT セクターのみならず，民間セクターにおける透明性規制のほとんどは，クレジット・スコアリングの場合にそうであるように（本章【16】を参照），全体的ないし大まかな情報へのアクセスを与えることを認めるだけである。他方，一般に情報自由法の適用を受ける公務員については，透明性を求める市民の利益にかんがみて，秘密は正当化されることが必要である。それにもかかわらず，各国の情報自由法は，法執行上の理由がある場合や，税法，競争法ないし金融市場規制の領域など，一定の状況下においては，公務員も戦略的に行動する必要があることを認めている。これらの状況において，政府が用いるデータベースやアルゴリズムに市民がアクセスできてしまうとすれば，法執行のための努力を阻害するか，あるいは実力あるアクターがシステムを「もてあそぶ」という結果になるかもしれない。このような事情から，ドイツの租税通則法（AO）88 条 5 項は，その公開が「徴税の平等性および適法性を危険にさらす」場合には，租税当局のリスクマネジメント・システムに関する情報へのアクセスを拒否できる旨を，明示的に規定している(40)。

【19】信託上の秘密の保護は，社会的相互作用において，当事者がしばしば，公衆および第三者に明かしたくない情報をやり取りしているという事実を考慮に入れたものである。秘密の情報のやり取りによって互いの協力が深められ，協調による戦略的行動が可能になり，信頼が安定化する。情報を秘密に保ちたいという意思は，関係性が内密であることに由来するかもしれず，あるいは，個人情報のやり取りのように，取り扱う情報の特別な性質に由来するかもしれない。後者はとりわけ，個人情報を取り扱う AI 搭載システムのデータベースについて当てはまる。この分野では，透明性を求める公共の利益と，プライバシー保護の必要性，個人情報の無欠性（integrity）の要請とが衝突する(41)。また，同様の衝突が生じるのは，私企業が排他的権利を有しているシステムを公的機関が利用している場合である。ここでは，公的な意思決定手続を「可視化し，理解可能に」しようとする利益が，システムの発展のために投資をしてきた民間事業者の権利とぶつかり合う（本章【17】を参照）。

【20】AI の秘密性に関する適法な主張といえども，包括的な例外取扱い（blanket exceptions）を正当化するものではない（GDPR の前文 63 段 6 文も参照）。〔上述の〕ドイツ国税通則法 88 条 5 項 4 文のように，政府の内部手続について外部的審査を完全に遮るような法規制は，厳格に過ぎる。また，クレジット・スコアリングの企業秘密に関してドイツ連邦通常裁判所（Bundesgerichtshof）はほとんど絶対的保護に近い立場を採っているが（本章【16】を参照），これは，低いスコアがユーザーに及ぼす実際上の影響が限定的である場合にのみ，正当化されうる。ここでは，互いに競合している透明性と秘密性の利益を調整することが重要であり，規制上の道具立てのなか

(39)「法的な秘密（legal secrets）」の理論につき参照，Scheppele (1988), Jestaedt (2001) and Wischmeyer (2018b).

(40) 参照，Braun Binder〔本書所収「AI と租税」〕, paras 12 et seq. さらに，参照，Martini and Nink (2017), p. 10.

(41) この問題は，あらゆる形態の透明性規制について当てはまる。参照，Holznagel (2012), § 24 para 74; von Lewinski (2014), pp. 8 et seq.

には，システムおよびその作動に関する重要な情報を提供しつつも，真にセンシティブな情報の保護を担保するような，様々な手段が存在する。そこに含まれるのは，⑴㋐通知義務と㋑一次データ／集合データ（aggregated data）へのアクセス権とを区別する，センシティブ情報のアクセスについての多段階の仕組み（multi-tiered access regimes）（本章【22】以下を参照），⑵アクセス権の一時的な制限，⑶情報媒介者（information intermediaries）を利用すること，⑷インカメラ手続などの手続的保護を採ること，である(42)。多くの場合，これらの手段を用いることで，システム運用者の秘密性の利益を失うことなく，全体的な透明性を高めることができる。しかし，透明性についての利益が秘密性に関する権利を上回り，合理的な調整を図ることができない場合には，その特定の場面において，AI 搭載システムを使用しないことについての重要な理由があるというべきである(43)。

2.3　機能的な制約

【21】一般的に，情報アクセス規制においては，多様なステークホルダーについてそれぞれに異なるアクセス権が認められている。例えば，報道機関は通常，情報自由法において一般的に認められる範囲よりも広範な法的権利を行使することができる。さらに，裁判所や公的機関が情報にアクセスしたり，それを生み出す権限は，報道機関のそれともまた異なってくる。このような細やかなアプローチは，関連するステークホルダーのそれぞれに特有の規範上の地位を考慮に入れるものであると同時に，当該情報が各ステークホルダーにいかなる意義を有するか，という機能的観点をも反映したものである。例えば，単なる市民としての一般的な利益に基づいて情報を得ようとするのではなく，自身が特定の意思決定により不利益を被るという理由から情報を求める場合には，アクセスを求める権利は通常，より強力なものとなる。このような仕組みを採用する他の理由として，複雑な問題をみずから評価する能力を必ずしも個々の市民が持っているわけではなく，そのような能力を獲得する必要もない，ということが挙げられる。むしろ個々の市民は，専門家ないし公的機関，とりわけ裁判所に頼ることが可能であり，頼るべきである。そうすることで，市民は自身の権利を主張することができ，かつ，より密度の濃い審査権を行使することができる(44)。

【22】このことが意味するのは，AI 規制に即していえば，画一的な解決手段を見つけようとするのはやめた方がよいということだ。すべての人が，あらゆる情報に対する無制限のアクセスを必要とするわけではない。AI 用の情報へのアクセスをデザインする際には，画一的な規制手法を採るのではなく，伝統的なアプローチにのっとって，次の諸点を考慮すべきである。すなわち，関連する個人の利益，各ステークホルダーの地位，公的機関・民間事業者などのシステム運用者の法的位置づけ，システムが用いられることになる領域のセンシティブ性・重大性といった事柄である（後述 4 を参照）。

【23】機能的アプローチを採用することで，AI の情報アクセス規制についてよくなされる批判に取り組むことが可能となる。それはすなわち，大多数の市民にとって，ソースコードへのアクセスを認めることが付加価値をもたらすことはないという批判である。批判者は他方で，システムに関する大まかな記述や，統計上重要な要素，全体的なアウトプットデータといった単純化された情報について市民のアクセスを認めたとしても，ただ明白なミスを発見することが可能になるだけであ

(42)　Wischmeyer (2018b), pp. 403–409.

(43)　これは，刑事司法ないし法執行の目的のために民間企業が所有するセンシティブな技術が用いられる場面を念頭に議論がされてきた。参照，Roth (2017), Imwinkelried (2017) and Wexler (2018). ここで述べた理由により，ノルトライン-ヴェストファーレン州の警察は，ニューラルネットワークを用いずに，決定木学習アルゴリズムによる予測的ポリシング（predictive policing）のシステムを開発した。参照，Knobloch (2018), p. 19.

(44)　ただし，（阻害的に働きうる）専門知のコストの問題もある。参照，Tischbirek〔本書所収論文〕, para 41.

り，アルゴリズムによる差別や情報保護違反といった真に重大な欠陥を見つけることはできないと主張する(45)。しかし，このことから必然的に，AIの透明性規制をすることは，アカウンタビリティの幻想を作り出すに過ぎず，状況を難しくするだけだと結論づけるべきだろうか（本章【3】参照）？このような批判をする者は，GDPR 13 条〜15 条に規定された権利義務などの情報に関する個人の権利にフォーカスし過ぎている。たしかに，情報アクセス権に関するこれまでの経験からすれば，個人の権利が万能薬でないことはわかる(46)。最悪のケースでは，そのような個人の権利の行使は，実際のポリシーを不透明なまま変化させない一方で，主観化を通じて社会問題を外部化させる（externalize）ことすらありうる。それにもかかわらず，透明性規制に関する一般的な議論のなかで，情報に対する個人の権利が民主主義社会において重大な役割を果たしうること自体を疑う者は，ほとんどいない。ただし，そのためには当該権利が，情報フロー規制の全体構造のなかの一要素として理解されることが必要である。またそれには，とりわけ規制当局ないし裁判所を通じてシステムの審査をすることで（本章【4】参照），得られた情報をより有意義に個人が活用できるようにする規制が伴わなければならない。

【24】しかし，情報へのアクセスという観点からすれば，個々の市民の役割は第一次的に道具的なものである。たしかに，市民がみずからの権利を行使することで，AI に関する社会の知識が増えることになる。しかし，多くの論者が指摘するように，透明性の目指すものは実際にはそれ以上に野心的である。もし，透明性が AI アカウンタビリティの基礎だと考えられるのであれば，「寄せ集めに含まれるありとあらゆる個々の要素を見ようとするのではなく，それがシステム全体としてどのように機能しているかを理解する」(47) ことが必要である。情報アクセス規制は，ステークホルダーにブラックボックスの中身を見させようとするが，それは必ずしも当該情報を理解可能にすることと結びつくものではない。情報を見ることとそれを理解することは同義だとする誤謬を避けるためには，〈AI 搭載システムの利用によって影響を受ける個人〉という視点に特にフォーカスした別の手段によって，情報アクセスの枠組みを補完する必要がある(48)。

3　説明を通じた行為主体性（エージェンシー）の創造

【25】もし AI 搭載システムにより決定・提案される借入の拒絶・税務監査・検索序列によって，個人が苦しめられ，または客体化されている(49)と感じるとすれば，ソースコードへの無制限のアクセスを認めたとしても，当該技術に対する彼らの信頼を回復することはほとんどできないだろう。それゆえ透明性規制は，情報を与えることだけに限定されてはならず，むしろ，問題となっている決定により影響を受ける者に行為主体性（エージェンシー）の感覚を与えるものでなければならないということが，広く認められてきている。このことが意味するのはつまり，影響を受ける人々が受け取る情報は，それに不服を申し立てる（脱出〔exit〕），ふるまいを変える（誠実〔loyalty〕），あるいは，民主主義社会において受け入れられるべき基準かどうかの議論を発議する（声〔voice〕）といった方法によって，有意義なかたちで何らかの決定に反応することを可能にしなければならない，ということである。同様の目的のため，GDPR 12 条 1 項およびその他の法律は，データ運用者に「簡素平明でわ

（45）　Datta et al. (2017), pp. 71 et seq.; Tene and Polonetsky (2013), pp. 269–270.
（46）　情報アクセス規制の強み・弱みに関する精緻な説明として参照，Fenster (2017).
（47）　Ananny and Crawford (2018), p. 983.
（48）　Ananny and Crawford (2018), p. 982.
（49）　GDPR 22 条の目的は，しばしば「影響を受ける当事者の人格や具体的事案の特性をまったく考慮せず，個人をして公権力による処理行為の単なる客体」へと格下げすること（degradation）を妨げるものとして定義される（Martini and Nink (2017), p. 3）。同旨のものとして，von Lewinski (2014), p. 16.

かりやすく，アクセスしやすい形式で，明瞭で平易な言語を用いた」データ主体への情報提供を要求する。AI 透明性をめぐる議論は，情報へのアクセスにあまりフォーカスし過ぎず，むしろ，AI 搭載システムの「明瞭性」「わかりやすさ」「理解可能性」「予測可能性」「説明可能性」を強化する方法を考えるべきだと主張されることがあるが，その論者の出発点にも上記と同様の考え方がある(50)。

【26】しかしながら，用語や概念が増加している事実は，情報を通じて個人の行為主体性（エージェンシー）がどのように作り出されうるかが完全には明らかではないことを示している。多くの人々にとって，説明可能性 ── この用語は，ここでは透明性を通じた行為主体性（エージェンシー）の感覚の創出という挑戦を象徴づけるものとして用いる ── は，因果関係と結び付けられている。したがって，説明可能 AI に関するコンピューター科学者によるごく最近の業績は，AI の基礎決定を生んでいる要因を特定し，アクセスしやすい方法でそれらの要因を伝えることに向けられている（**3.1**）。しかしながら，このアプローチには当然の限界がある。多くの事例において因果関係による説明（causal explanations）は，そもそも不可能であるか，あるいは，その影響を受ける者にとって容易にアクセスできないほどに複雑だからだ。そのため，近年では，因果関係による説明を代替する理論が，AI 透明性に関する学説において再び現れてきており，これは特にカウンターファクチュアルに関するデイヴィッド・ルイス（David Lewis）の業績に見られる（**3.2**）(51)。しかしながら，カウンターファクチュアルもまた，あらゆる説明において達成すべき，説明可能性と正確性（fidelity）のトレードオフから抜け出すことはできない。それゆえ，本章では視点を変えることを提案する。すなわち説明という行為を，個人への情報提供としてだけではなく，特定の制度的な仕組み（a specific institutional setting）のなかに埋め込まれていて，かつ，諸々の社会的・法的制度のなかで，あるいはそれらを通じて個人の行為主体性（エージェンシー）を生み出す社会的実践としても理解するように提案する（**3.3**）。

3.1　因果関係による説明

【27】科学哲学において「説明」とは，主として，ある事象または決定の原因となっている要因を特定することとして理解される(52)。倫理的な観点からいえば，因果関係による説明こそが行為主体性（エージェンシー）をもたらすといえる。というのも，因果関係による説明が存在することで，個人は，どうすればみずからの行為を修正することができ，また，結果を変えるためにはどの要素を問題にすべきかを認識しうるためである。人がなぜ特定の分類をされたかを学ぶことはすなわち，将来そのような分類をされることを避ける方法を知ることを意味する，と考えられている。この意味で，AI の説明可能性の強化とは，自分自身の作動を監視・分析し，因果的決定要因を特定し，人間に理解できる方法で影響を受ける者に対してそれを提示する能力を伴う AI 搭載システムを可能とする分析的ツールを開発することを意味する。多数のプロジェクトが，説明可能 AI の領域において現在進められている(53)。法的規制は，そのようなツール

(50)　参照，前掲注(13)および Doshi-Velez and Kortz (2017), p. 6。「説明は透明性とは区別される。〔この文脈で〕説明とは，AI システム内のビットの流れ（the flow of bits）を知ることを求めるものではない。このことは，人に説明を求める際に，ニューロンを通る信号の流れを知ることを求めるわけではないのと同じである」。

(51)　特に Wachter et al. (2018) が描写するルイスの業績，特に Lewis (1973a, b).

(52)　「科学的な説明，そして日常生活上のすべての説明が因果的なものであるかどうかという点についての哲学者たちの中での重要な反対意見，また（因果的な説明がたとえあるとしても）因果的な説明と非因果的な説明の間の区別についての反対意見」が存在する一方で，「実質的には誰もが（中略）原因についての情報を多くの科学的説明が引用するということに同意することに（中略）本質がある」(Woodward 2017)。さらに参照，Doshi-Velez and Kortz (2017), p. 3。

(53)　Cf. Russell et al. (2015); Datta et al. (2017), pp. 71 et seq.; Doshi-Velez and Kim (2017); Fong and Vedaldi (2018). 近年の発展にも関わらず，同領域における研究はいまだ初期段階である。2017 年には，XAI

の使用を義務付ける可能性がある。

【28】一部の学者は，GDPR 22 条 3 項がすでにそのような義務を含んでいると主張している(54)。本規定は ADMs の運用者に，GDPR 22 条 2 項(a)および(c)により自動処理プロセスが認められる場面においても「データ主体の権利を守るための適切な手段」を採るよう要求する(55)。しかしながら本規定を忠実に読むと，1995 年の EU データ保護指令の例にならって作られた GDPR 22 条は，個別事例において決定の根拠となった要素に関する情報を提供するよう ADMs の運用者に要求しているのではない。代わりに，利用されたシステムの論理（logic）に関する抽象的な情報を提供するだけで十分だとされている（本章【20】参照）。同じことが，GDPR 13 条 2 項(f)および 14 条 2 項(g)の通知義務にも，また，15 条 1 項(h)の情報にアクセスする権利にもあてはまる。法案の段階ではより野心的な「説明に対する権利（right to explanation）」についての議論がなされていたが，これは規則の最終版には容れられなかった(56)。

【29】もし，GDPR その他の法律がこのような因果関係による説明に対する権利を認めると仮定して，果たしてそれが技術的に実現可能かどうか，また，個人がその権利により利益を得るかを考えることは有意義であろう。この問いに対しては，上述の理由から（本章【10】以下参照），疑問が呈されている。特に，一部の AI 搭載システムについて因果関係の特定の困難性が指摘されている。たしかに，この点で説明可能 AI に関する近時の研究は発展を遂げてきている(57)。しかしながら，深層ニューラルネットワーク（deep neural networks）について，与えられるインプット変数と特定のアウトプットのあいだにある因果関係の特定を可能とするツールは，いまだ存在しない。さらに，大多数の AI 搭載システムは非常に複雑なので，そこでは，決定に導く因果的要因を，素人に理解できる方法で提示することが不可能である。意思決定を行う際にシステムが考慮する要因が複雑になればなるほど ── そして，その決定数の増加は，通常，人間の知能に代わって AI を用いる真の目的である(58)──，その要因にアクセスするのはますます困難になる。これは AI についての真実であるのみならず，すべての意思決定システムの特徴である。説明可能 AI のさらなるイノベーションを想定すると，深層学習システムは，最終的には意思決定プロセスに影響しているすべての要因を列挙することを可能にし，さらにはその要因を，個別事案における統計的な重要性に応じてランク付けすることまで，可能になるかもしれない。しかし，あらゆる相互関係とともに列挙される機能（features）の総数が，情報処理についての個人のキャパシティを超えるならば，その列挙されたリストは，コードそのものと同様，名宛人にとって意味のないものとなるだろう(59)。

【30】この問題に取り組むために，機械学習コミュニティから多様でイノベーティブな提案がなされてきた。そこには例えば，LIME（Locally Interpretable Model Agnostic Explanations）や，ベイズ決定のリスト（Bayesian Decision Lists（BDL））， BETA （Black Box Explana-

についての DARPA プロジェクトがはじまった（参照 www.darpa.mil/program/explainable-artificial-intelli gence）。

(54)　Goodman and Flaxman (2016). 追加的に，Wachter et al. (2017), pp. 76–77 も参照のこと。

(55)　この規定の限定範囲について，参照，前掲注(33)。

(56)　立法過程の詳細な分析については，参照，Wachter et al. (2017), p. 81; Wischmeyer (2018a), pp. 49–52。

(57)　例えば，学者らは近年，画像を分類するために用いられる深層ニューラルネットワークの分類子間の命令関係を構築するメカニズムを提案しており，同技術についての因果的モデルはめざましい発展を遂げている（Palacio et al. (2018)）。さらに参照，Montavonet al. (2018)。

(58)　参照，Hermstrüwer〔本書所収「AI と不確実性のもとでの行政決定」〕, paras 70–74.

(59)　Ribeiro et al. (2016), sec. 2 は次のように述べる：「数百，あるいは数千の特徴が予測（prediction）に関わるとしたら，仮に個々の重要性を調査しうるとしても，なぜその予測ができるのかをユーザーが理解するよう求めることは合理的ではない」。

tions through Transparent Approximations）などが含まれる。これらを用いることで，当該モデルが信頼しうるかを知るために，機械学習システムにおける分類子（classifiers）の信頼性をテストすることが可能となる。現在のところ，これらのモデルは，ある技術について専門家が詳しく理解しようとする際に役立てられている。他方，これらのモデルが素人にも役立つかどうかは，いまだ開かれた問題である(60)。いずれにせよ，このアプローチにおいて私たちは，一方では AI に基づく意思決定プロセスの複雑性を素人にも理解できる水準にまで低減しつつ，他方で同時に，その説明の価値の有益性を保とうとする，その 2 つの目標のあいだに内在的な緊張があるという事実に，再び直面する。つまるところ技術は，正確性と解釈可能性のトレードオフを何とか最適化しようと試みるが，ここから逃れることはできないのだ(61)。

3.2　カウンターファクチュアルな説明

【31】意思決定プロセスの厳密な因果的分析のみをもって「説明」とみなすならば，正確性と解釈可能性のトレードオフは，〈説明を通じた行為主体性（エージェンシー）の創出〉というプロジェクトそれ自体を脅かしかねない。しかしながら，「特定の決定に関する」あらゆる「本当の決定要素（real determinants）(62)」を特定することは，たしかに説明可能性を支えるかもしれないが，それは，説明〔なる行為〕が一般的に理解され，社会的実践として運用されている方法にとっての，必要条件でも十分条件でもない。現代社会は，認識的な観点からみれば，常に「不透明性の交響曲（symphony of opacity）」(63)（ニクラス・ルーマン〔Niklas Luhmann〕）であり続けてきた。それなのに，もし仮に，説明可能性のためにはあらゆる決定要素の特定が必要だというのであれば，およそ，この世の中で説明を求める意見は，ほとんど意味をなさないだろう。特に，（人間による）集合的決定（collectve（human）decisions）というのは，構造上，説明不可能なものである。なぜなら，多くの場合，「真の」動機となる原因を可視化するようなかたちで当該決定を再現することは不可能だからである(64)。また自らの個人的な決定についても，多くの場合，本当の原因（true causes）を追い求めてみたところで，結論までたどり着くわけではない(65)。

【32】それにも関わらず，私たちはいまだに，個人も集団も，みずからの行動を説明すべきだと主張する。法システム内部では，行為主体性（エージェンシー）の感覚を市民に持たせるため，公権力の大多数の決定について，説明が義務的なものとされている（本章【37】参照）。これが示唆するのは次のことである。すなわち，日々の実践に根付いており，かつ，社会で実際になされている行為としての〈説明〉の概念は，科学的に原因を突き止める行為からは，概念的に区別されるということである(66)。この洞察は，説明を通じて AI 透明性を創りだすための代替的手段を発見しようとしている学者 ── その数はますます増えている ── のコミュニティにお

(60)　Wachter et al.（2018）, p. 851.
(61)　このトレードオフ関係について，参照，Lakkaraju et al.（2013）, sec. 3; Ribeiro et al.（2016）, sec 3.2。Wachter et al.（2018）, p. 851 は「近似値（approximation）の質と，機能の理解の容易性，および，当該近似値が妥当となる領域の広さの 3 点がトレードオフの関係」にあるとさえ述べる。
(62)　Bundesverfassungsgericht 2 BvR 1444/00（20 February 2001）, BVerfGE 103, pp. 159–160.
(63)　Luhmann（2017）, p. 96.
(64)　Wischmeyer（2015）, pp. 957 et seq.
(65)　Tutt（2017）, p. 103.　Lem（2013）, pp. 98–99 は次のように述べる：「すべての一人ひとりの人類が，アルゴリズムを知らずに用いることができるデバイスの，すばらしい実例である。私たちの脳こそが，全宇宙の中で『私たちに最も近い（closest to us）』『デバイス』のひとつである。なぜなら，私たちはそれを自分の頭の中に持っているからだ。現在でもまだ，私たちは脳がどのように動くのか，正確には知らない。心理学の歴史が示すように，自らを省みることによってその仕組みを知ろうとする行為は，大いに間違いを含み，人を誤らせ，いくつかの極めて間違った仮説へと連れていく」。
(66)　人間は，他者の決定を生み出す原因を正確に知らない場合にさえ，相互に作用しあうことができる。この事実は，ひょっとすると〔生物的〕進化のなかの一要素なのかもしれない（Yudkowsky（2008）, pp. 308 et seq.）。

いて，共有されている。なかでも特に卓越した提案に，「仮定的変更（hypothetical alterations）」という見解がある。これはダニエル・シトロン（Danielle Citron）とフランク・パスカル（Frank Pasquale）がクレジット・スコアリングの議論に用いる見解であるが，消費者について信用履歴（credit histories）へのアクセスを認め，かつ，（〔「もしも…」という〕仮定条件において）その履歴の修正を認めることによって，効果を分析するというものである(67)。

【33】同様に，ザンドラ・ヴァクター（Sandra Wachter），ブレント・ミッテルシュタット（Brent Mittelstadt），クリス・ラッセル（Chris Russell）は最近，「カウンターファクチュアルな〔反実仮想の〕説明」という見解を取り入れた。これによれば，AI に基づく意思決定プロセスの要因のうち，異なる結果に到達しようとする場合に変更をすることが必要となる要因に限って，開示が必要だとされる(68)。例えば融資申請がうまくいかなかったケースでは，個々の申請者は，仮に良い結果を受け取るためにはどのデータが変更・訂正される必要があったか，カウンターファクチュアルな説明を与えられる。彼らの主張によれば，そのような情報は，申請者が自身の行動を修正し，不利益な決定を争うことを可能にする(69)。ここでは，仮定的提案（the hypothetical）によって具体的な代替的未来が示されるため，その動機づけの力は，「単なる」因果関係による説明の場合と比較して，実際に強くなる可能性がある。ヴァクターらによれば，システム運用者は，決定が下されたとき，もしくはその後のタイミングで，自動的に，あるいは，個人による特定の要求に応じるかたちで，カウンターファクチュアルを算定し，それを開示するよう義務付けられうる(70)。

【34】カウンターファクチュアルの概念は魅力的だ。なぜなら，この考え方は，社会のなかで，意思決定プロセスを厳密に再現することを，説明において求めることはほぼできないという洞察に基づいているためである。代わりに説明は，通常，決定の相違を生み出すような事実に焦点を限定する。さらに，この理論は，AI 搭載システムが素人にとっての「ブラックボックス」であり続けることを，ある程度は必然的に認めることになるだろう（本章【10】参照）。また，この理論では，外部からシステムの作動を監視するだけなので，透明性と秘密性の利益の均衡をとる手続きを回避することができる。

【35】しかしながら，他の重大な限定に加えて(71)，カウンターファクチュアルは正確性と解釈能力のトレードオフから逃れていないことが指摘されうる(72)。ヴァクターらが認めている通り，データ主体が異なる結果を得ようとする場合に調整しなければならない要素を説明するためには，意思形成システムの複雑性が高まれば高まるほど，ますます多くのカウンターファクチュアルが必要とされる(73)。したがって，カウンターファクチュアルは，ほどほどに複雑な AI 搭載システムの一般的な理解を深めようとするデータ主体を助けることは得意である。しかし，主としてそれによって利益を得るのは，データ主体や第三者に対して深

(67)　Citron and Pasquale (2014).
(68)　Wachter et al. (2018). Doshi-Velez and Kortz (2017), p. 7 も参照のこと。哲学的根拠については，参照，Lewis (1973a, b) and Salmon (1994)。
(69)　Wachter et al. (2018), p. 843.
(70)　Wachter et al. (2018), p. 881.
(71)　Wachter et al. (2018), p. 883 は，次のように述べる：「説明の最小限度の形態であるため，カウンターファクチュアルは，あらゆるシナリオに適合するわけではない。特にシステムを機能的に理解し，または，自動的処理決定の理論的根拠を理解することが重要な場面では，統計的なエビデンスを提供しないカウンターファクチュアルは，公正性や人種バイアスに関わるアルゴリズムの評価を必要とする」。
(72)　同旨のものとして，Hermstrüwer〔本書所収「AI と不確実性のもとでの行政決定」〕，paras 45–48.
(73)　Wachter et al. (2018), p. 851 は，次のように述べる：「欠点は，個々のカウンターファクチュアルが過度に制限的となりうることである。すなわち，単一のカウンターファクチュアルが，未来の決定の前にデータ主体によって変更されえず，かつ，正確なデータに，いかにして決定が基礎づけられているかを示すことがありう

層アクセス（deep access）を与えることなく，システムについての情報を提供することができる，そのようなシステムの運用者たちなのだ。

3.3　文脈上の説明

【36】説明が，決定に関する情報が名宛人に示される際の特別な形態（a specific form）としてしか理解されないのであれば，正確性と解釈能力のトレードオフ関係は避けることができない。しかし，AIの「説明可能性」のほとんどすべての理論で欠けているのは，説明が通常行われる際の制度的な仕組みである。この仕組みは重要なものだ。なぜなら，〔行政機関・裁判所などの〕制度（インスティテューション）＝機関が，個人に任せられている解釈の負担の一部を取り除くことができるからであり，それによって，解釈可能性を維持するために正確性について妥協をする，という必要が少なくなるためである。

【37】法システムにおいて，制度（インスティテューショナル）＝機関的観点は，説明 ── より明確に言えば理由提示（reason-giving）── の理論を構築するうえで必要不可欠なものである。理由提示の目的の法的な必要条件は，決定について個人に通知することだけでなく，彼らの権利を保障するため，法の支配の複雑な仕組みを使用可能にすることでもある(74)。ドイツ連邦行政裁判所は，理由提示と司法へのアクセスの結びつきを強調し，次のように述べる：「自身の権利を十分に擁護できるよう，行政活動によって自身の権利に影響を被る市民は，関連する重要な理由を告げられなければならない」(75)。

【38】ここにいう説明を，制度的文脈に埋め込まれた社会的実践として理解することは ── もちろん説明は，協働的な統制プロセスのほんの初めの段階に過ぎないのだが ──，因果関係についての情報およびカウンターファクチュアルな情報を提供することの重要性を否定するものではない。むしろ，それは視点を転換させるものである。すなわち，行為主体性（エージェンシー）は，ある特定の決定やありうる代替案の原因（causes）に関する知識を通じて得られるだけではない。むしろ，当該決定を解釈し精査する個人をサポートする〔行政機関・裁判所などの〕制度（インスティテューション）＝機関によっても行為主体性（エージェンシー）が生み出されうるという事実に，注意してほしい。以下に述べる通り，この制度的アプローチは，あらゆる説明のなかに存在する2つの異なる要素を本質的に区別することによって，正確性と解釈可能性のジレンマから，法システムが抜け出すことを可能とする。

【39】ここにいう説明の2つの要素のうち，第一は，名宛人たる市民に対して次の認識を確実にさせるものである。すなわち，当該名宛人がある決定に服従させられたこと，そしてその際，その結果を受け入れるか，または，〔不服申立ての審査庁たる〕上級行政庁ないし裁判所を巻き込んで法的救済を求めていくのか，いずれかを選択することができる，という事実を認識できるようにすることである。もし名宛人が説明を完全に理解しない場合には，法的アドバイスを得ようとすることが予想される(76)。これに対して，第二の要素は〔システムを〕監督する公的機関に対して向けられている。〔素人の市民とは異なり，裁判所・行政機関などの〕監督機関についていえば，説明は，効果的な統制を行いうるために必要な限りで，複

る。それも，有利な結果のためには修正されねばならない他のデータが存在する可能性があるにもかかわらず，である。この問題は，複合的で多様な事実に反する説明がデータ主体に提示されることによって，解決されうる」。

（74）（因果的な）説明と（意味論的な）理由は同一ではない。しかし，説明と理由提示の両者が効果的であるためには，いずれも制度的な枠組みを必要とする。理由提示の要求は，第一に公的権力のために存在する。しかし，裁判所や学者によって記述される理由提示の機能は，私的な集団にとっても効果がある。より詳細には，参照，Kischel (2003), pp. 88 et seq.; Wischmeyer (2018a), pp. 54 以下。比較的な分析として，参照，Saurer (2009), pp. 382–383。

（75） Bundesverwaltungsgericht 2 C 42.79 (7 May 1981), DVBl 1982, pp. 198–199.

（76） Stelkens (2018), § 39 VwVfG, paras 41, 43.

雑なものでなければならない。したがって，ある説明が素人にとってどれほど分かりやすいもの〔であるべき〕かという問題は，それにより影響を受ける個人のキャパシティに左右されるのではなく，むしろ，各事案の状況に左右されるのである。この点をニクラス・ルーマンは，冷静かつ皮肉っぽい彼らしいやり方で，次のように述べている：

> したがって実際には，説明の文章というのは，法律家によって，法律家のために書かれる。あるいは，上級行政庁，裁判所のために書かれる。その際，正しく，間違いのないようにという努力が行き過ぎて，その文章はしばしば暗号化（エンコード）されてしまう。あまりに複雑なので，情報の受け手はその文章を自分では理解できず，専門家の助けを借りてやっと解読（デサイファー）できるのである(77)。

【40】このモデルにおいて，説明が成功するか否かは，次の 2 点にかかっている。それは第一に，自らが決定に服従していることを解読（デサイファー）する個人のキャパシティであり，第二に，〔行政機関・裁判所など，システムを〕統制する側の制度＝機関（インスティテューション）が有している専門的知識の完全性とその水準である。現在，AI についてこのモデルを採用するためには，深刻な障害が存在している。というのも，現在のところ，法的統制を担う制度＝機関（インスティテューション）は，多くの複雑な問題を取り扱う能力を有する一方で，AI を取り扱う能力を有していないからだ。現在のところ，行政機関も裁判所も，複雑な AI 搭載システムを分析する経験をほんのわずかしか有していない。そのため，そのキャパシティの発展と強化こそが，この領域におけるすべての真摯な政治的戦略と同様，AI の透明性に関するあらゆる理論の必要不可欠な部分をなすことは間違いない。

4　今後の道筋

4.1　基本的方針

【41】AI 透明性についての議論では，間違った絶対者に打ち勝つことが必要だ。すべての AI 搭載システムが不可解なブラックボックスなのではないし，透明性が，それ自身によってアカウンタビリティを完全に保障するのでもない(78)。むしろ AI 透明性規制の価値は，次の点にある。すなわち，知識を生み出すことで技術についての議論を引き起こすこと，AI に基づく決定に個人が異議を申し立てるよう動機を与えること，そして ── 長期的にみて，最終的には ── 新しい技術の社会的な受容を強化することである(79)。規制がどの程度実際にこれらの目標に到達できるかは，様々な要因に左右される。しかし，次のことは少なくとも妥当だといえそうである。すなわち，第一に，大規模な協働の統制プロセスのなかに埋め込まれた，より強固な情報アクセス規制の存在と説明提示（explanation-giving）義務の導入が，上記の目標を実現することに資するであろうということ。第二に，情報の提供および決定の説明についての拒絶が，抵抗を引き起こし，信頼を失わせるかもしれないということ(80)。したがって透明性規定は，広範囲にわたる無知（ignorance）に抗い，そして AI の台頭に伴って生じる権利剥奪（disfranchisement）の感覚が広まるのを阻止する手段とみなされるべきである。

【42】このアーキテクチャの将来設計にとって重要なのは，〈情報へのアクセス〉と〈説明〉の区別である。この点は，これまでの各章で丁寧に議論してきた。そして，同様に重要なのは，AI 搭載システムの運用者が公的機関か民間事業者か，という性質の差異である。この点は，AI の透明

(77)　Luhmann (1983), p. 215.
(78)　透明性をその一局面とする AI アカウンタビリティの問題についての包括的な議論について，参照，Busch (2018)。
(79)　この最後の点は，特に EU およびアメリカの憲法において重要である。Saurer (2009), pp. 365 and 385 を参照。
(80)　Mittelstadt et al. (2016), p. 7.

性についての包括的な理論において，さらに深められる必要があるだろう（本章【17】以下参照）。加えて，AI の透明性と秘密性という競合する利益のバランスをとる際に，規制当局は次の要素に注意する必要がある ── ⑴システムが採用されている分野の重要性，⑵当該分野のシステム運用者の相対的重要性，⑶システムの展開によって影響を受ける個人および集団の権利の質と重要性，⑷関係するデータの種類（特に GDPR 9 条の意味での個人データの特定のカテゴリに関心があるかどうか），⑸当該技術が個別の意思決定プロセスに関与する程度 ── なお，AI 透明性に関する法は，GDPR 22 条の形式的なアプローチ（それは自動処理プロセスがデータ主体に対して「法的効果」ないし「同様に重要な影響」を有することを要件とする）を克服し，むしろ，個人情報を扱っており，かつ，意思決定プロセスに資するようなあらゆる AI 搭載システムを対象とすべきである[81]──，⑹組織的・技術的・法的な手段によって，AI の秘密性に関してシステム運用者が有している正当な利益を保護することが可能かどうか，また，どのようにしてそれが可能となるか。

4.2 実　践

【43】透明性はアカウンタビリティの情報的側面であり，その意味で，意思決定プロセスの統制枠組みの前提条件であり，かつ，その産物でもある。それゆえ，透明性規制は，多面的な問題である。「アナログ」な世界では，秘密保護法（secret laws）の禁止，行政および司法の活動に対する手続的透明性の要求，議員・プレス・市民の情報に関わる異なる憲法上・制定法上の権利，といった様々な手段が，公的アクターの決定の透明性を確保するために用いられている。このことは，AI 透明性の規制についても同じく妥当する。情報および説明について個人の権利は，より大きな規制的・監督的構造の内部にある一つの要素にすぎない。技術 ── すなわち，デザインによる透明性（transparency by design）── は，このアーキテクチャのもう一つの重要なパーツであるが，しかし，それが唯一の答えでもない。このような背景のもと，有意義な透明性の戦略は，少なくとも以下の 5 つの相互に連動的な手段を用いるべきである。

【44】第一の，そしておそらく最も論争的でないステップは，AI 搭載システムの存在を開示する義務を課すことであり，それと同時に，権限ある行政機関および最終的にはデータ主体に対して，その使用について通知する義務を課すことであろう[82]。いくつかの分野では，そのような開示と通知の要求は，すでに存在するか，あるいは現在議論がされている[83]。公的機関の場合，開示する範囲のなかに「その目的，範囲，可能性のある内部利用の方針あるいは内部利用の実務，そして実施のタイムラインについての詳細」が含まれる[84]。ここでは通知と開示が法の支配の基本的な要求であるため，例外は，特別な状況に限って正当化されうる[85]。これに対して，民間のアクターにとっては，とりわけ，システムが用いられる領域や，影響を受けるデータ主体の権利の質・強度，上述のような〔透明性の〕要求を法定できるかどうかに左右される。しかしながら，上述のような抽象的に示された情報がとりわけセンシティブだということはほとんどないため，立法者には一定の裁量が認められる。

【45】AI 搭載システムが個人の意思決定の目的に役立てられるとしたら，通知は説明を求める権利と結び付けられる可能性があり，これが透明性アーキテクチャの第二のレベルである。**3** で述べたように，説明とは，名宛人に対して，システム

(81)　参照，前掲注(34)。同様の主張として Wachter et al. (2018), p. 881.
(82)　Martini (2017), p. 1020; Busch (2018), pp. 58–59.
(83)　参照，前掲注(4)〜注(9)。
(84)　Reisman et al. (2018), p. 9.
(85)　参照，Bundesverfassungsgericht 1 BvR 256, 263, 586/08 'Vorratsdatenspeicherung' (2 March 2010), BVerfGE 125, pp. 336–337.

の正確な記述を与えることや，あらゆる原因やカウンターファクチュアルのリストを提供することではない。むしろ，説明において提示される情報については，〔システムを〕統制する実務において裁判所や上級行政庁をサポートすることや，決定に対して行政的・司法的統制を求める権利を市民に実現させることなど，説明提供義務の目的の観点から，それがどこまで必要か，注意深く測定される必要がある。これには次のような情報が含まれるかもしれない：⑴システムのデータ基礎，⑵モデルと決定論理（decision logic），⑶システム運用者の（データ）品質標準，⑷システムで用いられる参照グループまたはプロファイル，⑸関係する個人に関するシステムの実際のまたは潜在的な推論，など[86]。そのような説明の正確な内容は，抽象的に定義されることはありえず，個別ケースでそれぞれ決定されねばならない[87]。

【46】第三の段階は，情報にアクセスする個人の権利である。そのような権利は，データ主体あるいは研究者に対して，もし仮に彼ら自身が AI 搭載システムの決定によって影響を受けていない場合であっても，データ運用者に対する司法あるいは行政の活動を準備し，あるいは単にシステムについて議論を引き起こすために用いられうるシステムの評価とその実施を認める。公的機関の場合，そのような権利を生み出すための自然な方法は，情報自由法の適用範囲を拡張し，AI 搭載システムをそこに含めることである[88]。民間のアクターが情報にアクセスする権利については，いまだに未成熟で研究が十分でない問題だ。**2.2** で議論したように，情報にアクセスする権利は完全ではない。現在の情報自由法は，AI の秘密性にも適用されるであろう例外を含んでいる。それは例えば，行政が空港セキュリティや脱税捜査のためにシステムを利用するような場合である。民間のアクターに関していえば，秘密性への権利は，情報にアクセスする権利が競業他社によって不正に用いられるのを阻止しようとする際に，典型的に強化される。ここでは競合する利益の間の合理的な均衡が，アクセス権の例外の慎重な設計を通じて達成されなければならない（本章【20】参照）。

【47】それらの例外によって情報に関する権利の実践的効果が削がれる可能性があるものの，システムのアカウンタビリティはその他の手段によっても保障されうる。特に監督機関や裁判所も，情報にアクセスする権利を与えられるかもしれない。このように，公的機関に調査権限を付与することは，包括的な AI 透明性レジームの第四の段階である。民間の領域においては，公的な調査権限は，行政的統制の標準的な手段である。そして公的機関自身が AI 搭載システムを用いるときには，より上級のまたは専門的な監督官庁または裁判所が，決定の適法性を審査するため，システムにアクセスできることが必要である（参照，ドイツ基本法 19 条 4 項）。秘密性の利益にも，この文脈で注意を払う必要があるが，ここでは通常，個人の情報に対する権利の場合に比べて問題は容易におさめられうる。なぜなら，情報は国家の領域に留まるからである。監督メカニズムがどれほど厳密であるべきかは，**4.1** で議論した要因に依存する。特に，個人の情報アクセス権が与えられないところでは，監督的統制は埋め合わせ機能を有し，それゆえ徹底的に，かつ，より頻繁に行われなければならない。AI 搭載システムの調査権限が一つの専門的な機関に集中的に与えられるのか[89]，または様々な機関がそれぞれの管轄内で AI 搭載システムの使用を監督しうるように権限を割り振るかどうかは，各国の行政文化や憲法上の要求にも左

（86） 参照，Busch（2018），pp. 59–60; Sachverständigenrat für Verbraucherfragen（2018），pp. 122, 162–163; Zweig（2019）.

（87） 参照，前掲注(38)。

（88） ある学者によれば，このような要求はフランスのデジタル共和国法を通じてすでに導入されているという（Loi n° 2016–1321 pour une République numérique）。同法は，「コードソース」という用語によって，公衆行政関係法典（Code des relations entre le public et l'administration）の 300–2 条における行政文書の定義を変更したものである。詳細については参照，Jean and Kassem（2018），p. 15。

（89） 参照，Tutt（2017）。

右される（参照，ホフマン-リーム・本書所収「法と規制に対する挑戦としての AI」39 段落）。技術的な用語を用いるならば，AI 搭載システムの運用者は，これらの監督機関のための適切な技術インターフェイス（APIs）を作る必要があるだろう。

【48】第五の，そして最後の段階は，AI 搭載システムの作動を記録する義務のような，様々な付帯的手段を含む（本章【16】参照）[90]。さらに，裁判所や諸機関は，AI 搭載システムを統制するための高度な専門的知識を必要な水準で用意しなければならない。そのためには，政府は専門家ネットワークに参加し，専門的機関と標準化プロセスを立ち上げ，〔システムの〕認証と検査の枠組みを作るべきである[91]。言うまでもなくこれは，透明性規制の範囲を超える，困難で長期的なタスクであるが，「ブラックボックスをこじ開ける」ための真摯な努力として絶対不可欠なものなのだ。

参考文献

Agency for Digital Italy（2018）White Paper on artificial intelligence at the service of citizens. www.agid.gov.it/en/agenzia/stampa-e-comunicazione/notizie/2018/04/19/english-versionwhite-paper-artificial-intelligence-service-citizen-its-now-online

Ananny M, Crawford K（2018）Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. New Media Soc 20（3）: 973–989

Arbeitsgruppe “Digitaler Neustart”（2018）Zwischenbericht der Arbeitsgruppe “Digitaler Neustart” zur Frühjahrskonferenz der Justizministerinnen und Justizminister am 6. und 7. Juni 2018 in Eisenach. www.justiz.nrw.de/JM/schwerpunkte/digitaler_neustart/zt_fortsetzung_arbeitsgruppe_teil_2/2018-04-23-Zwischenbericht-F-Jumiko-2018%2D%2D-final.pdf

Article 29 Data Protection Working Party〔第 29 条作業部会〕（2018）Guidelines on automated individual decisionmaking and Profiling for the purposes of Regulation 2016/679（wp251 rev.01）. ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

Asilomar Conference（2017）Asilomar AI principles. futureoflife.org/ai-principles

Ben-Shahar O, Chilton A（2016）Simplification of privacy disclosures: an experimental test. J Legal Stud 45: S41–S67

Ben-Shahar O, Schneider C（2011）The failure of mandated disclosure. Univ Pa Law Rev 159: 647–749

Buchner B（2018）Artikel 22 DSGVO. In: Kühling J, Buchner B（eds）DS-GVO. BDSG, 2nd edn. C. H. Beck, München

Bundesanstalt für Finanzdienstleistungsaufsicht〔ドイツ連邦金融監督庁（通称 BaFin）〕（2018）Big Data trifft auf künstliche Intelligenz. Herausforderungen und Implikationen für Aufsicht und Regulierung von Finanzdienstleistungen. www.bafin.de/SharedDocs/Downloads/DE/dl_bdai_studie.html

Burrell J（2016）How the machine ‘thinks’: understanding opacity in machine learning algorithms. Big Data Soc 3: 205395171562251. https://doi.org/10.1177/2053951715622512

Busch C（2016）The future of pre-contractual information duties: from behavioural insights to big data. In: Twigg-Flesner C（ed）Research handbook on EU consumer and contract law. Edward Elgar, Cheltenham, pp. 221–240

Busch C（2018）Algorithmic Accountablity, Gutachten im Auftrag von abida, 2018. http://www.abida.de/sites/default/files/ABIDA%20Gutachten%20Algorithmic%20Account

（90） データ運用者は「構造的で，一般に用いられ，機械で読み取り可能なフォーマット」によって，個人あるいは監督機関にとって有用な記録されたデータを作る必要がある（参照，GDPR 20 条）。異なる文脈におけるこの要求について，参照，Bundesverfassungsgericht 1 BvR 1215/07 ‘Antiterrordatei’（24 April 2013）, BVerfGE 133, p. 370 para 215.

（91） 参照，Kaushal and Nolan（2015）; Scherer（2016）, pp. 353 以下；Martini and Nink（2017）, p. 12; Tutt（2017）, pp. 83 以下。一般的な政府の理解について，参照，Hoffmann-Riem（2014）, pp. 135 以下。

ability.pdf
Citron D, Pasquale F (2014) The scored society: due process for automated predictions. Washington Law Rev 89: 1–33
Costas J, Grey C (2016) Secrecy at work. The hidden architecture of organizational life. Stanford Business Books, Stanford
Crawford K (2016) Can an algorithm be agonistic? Ten scenes from life in calculated publics. Sci Technol Hum Values 41 (1): 77–92
Datenethikkommission (2018) Empfehlungen der Datenethikkommission für die Strategie Künstliche Intelligenz der Bundesregierung. www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/2018/empfehlungen-datenethikkommission.pdf?__blob=publicationFile&v=1
Datta A, Sen S, Zick Y (2017) Algorithmic transparency via quantitative input influence. In: Cerquitelli T, Quercia D, Pasquale F (eds) Transparent data mining for big and small data. Springer, Cham, pp. 71–94
Diakopoulos N (2016) Accountability in algorithmic decision making. Commun ACM 59 (2): 56–62
Doshi-Velez F, Kim B (2017) Towards a rigorous science of interpretable machine learning. Working Paper, March 2, 2017
Doshi-Velez F, Kortz M (2017) Accountability of AI under the law: the role of explanation. Working Paper, November 21, 2017
Edwards L, Veale M (2017) Slave to the algorithm? Why a 'Right to an Explanation' is probably not the remedy you are looking for. Duke Law Technol Rev 16 (1): 18–84
European Commission (2018) Artificial intelligence for Europe. COM (2018) 237 final
European Parliament (2017) Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics. 2015/2103 (INL)
Fassbender B (2006) Wissen als Grundlage staatlichen Handelns. In: Isensee J, Kirchhof P (eds) Handbuch des Staatsrechts, vol IV, 3rd edn. C. F. Müller, Heidelberg, § 76
Fenster M (2017) The transparency fix. Secrets, leaks, and uncontrollable Government information. Stanford University Press, Stanford
Florini A (2007) The right to know: transparency for an open world. Columbia University Press, New York
Fong R, Vedaldi A (2018) Interpretable explanations of Black Boxes by meaningful perturbation, last revised 10 Jan 2018. arxiv.org/abs/1704.03296
Goodman B, Flaxman S (2016) European Union regulations on algorithmic decision-making and a "right to explanation". arxiv.org/pdf/1606.08813.pdf
Gusy C (2012) Informationsbeziehungen zwischen Staat und Bürger. In: Hoffmann-Riem W, Schmidt-Aßmann E, Voßkuhle A (eds) Grundlagen des Verwaltungsrechts, vol 2, 2nd edn. C. H. Beck, München, § 23
Harhoff D, Heumann S, Jentzsch N, Lorenz P (2018) Eckpunkte einer nationalen Strategie für Künstliche Intelligenz. www.stiftung-nv.de/de/publikation/eckpunkte-einer-nationalenstrategie-fuer-kuenstliche-intelligenz
Heald D (2006) Varieties of transparency. Proc Br Acad 135: 25–43
Hildebrandt M (2011) Who needs stories if you can get the data? Philos Technol 24: 371–390
Hoffmann-Riem W (2014) Regulierungswissen in der Regulierung. In: Bora A, Reinhardt C, Henkel A (eds) Wissensregulierung und Regulierungswissen. Velbrück Wissenschaft, Weilerswist, pp 135–156
Hoffmann-RiemW (2017) Verhaltenssteuerung durch Algorithmen — Eine Herausforderung für das Recht. Archiv des öffentlichen Rechts 142: 1–42
Holznagel B (2012) Informationsbeziehungen in und zwischen Behörden. In: Hoffmann-Riem W, Schmidt-Aßmann E, Voßkuhle A (eds) Grundlagen des Verwaltungsrechts, vol 2, 2nd edn. C. H. Beck, München, § 24
Hood C, Heald D (2006) Transparency. The key to better Governance? Oxford University Press, Oxford
House of Lords Select Committee on Artificial Intelligence (2018) AI in the UK — Ready, willing and able? publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf
IBM (2018) Continuous relevancy training.

console.bluemix.net/docs/services/discovery/continuous-training.html#crt

Imwinkelried E (2017) Computer source code. DePaul Law Rev 66: 97–132

Jean B, Kassem L (2018) L'ouverture des données dans les Universités. openaccess.parisnanterre.fr/medias/fichier/e-tude-open-data-inno3_1519834765367-pdf

Jestaedt M (2001) Das Geheimnis im Staat der Öffentlichkeit. Was darf der Verfassungsstaat verbergen? Archiv des öffentlichen Rechts 126: 204–243

Kaushal M, Nolan S (2015) Understanding artificial intelligence. Brookings Institute, Washington, D. C. www.brookings.edu/blogs/techtank/posts/2015/04/14-understanding-artificial- intelli gence

Kaye D (2018) Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression 29 August 2018. United Nations A/73/348

Kischel U (2003) Die Begründung. Mohr Siebeck, Tübingen

Knobloch T (2018) Vor die Lage kommen: Predictive Policing in Deutschland, Stiftung Neue Verantwortung. www.stiftung-nv.de/sites/default/files/predictive.policing.pdf (19 Jan 2019)

Konferenz der Informationsfreiheitsbeauftragten〔情報自由監察官会議〕(2018) Positionspapier. www.datenschutzzentrum.de/uploads/informationsfreiheit/2018_Positionspapier-Transparenz-von-Algorithmen.pdf

Lakkaraju H, Caruana R, Kamar E, Leskovec J (2013) Interpretable & explorable approximations of black box models. arxiv.org/pdf/1707.01154.pdf

Leese M (2014) The new profiling: algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union. Secur Dialogue 45 (5): 494–511

Lem S (2013) Summa technologiae. University of Minnesota Press, Minneapolis

Lewis D (1973a) Counterfactuals. Harvard University Press, Cambridge

Lewis D (1973b) Causation. J Philos 70: 556–567

Luhmann N (1983) Legitimation durch Verfahren. Suhrkamp, Frankfurt am Main

Luhmann N (2017) Die Kontrolle von Intransparenz. Suhrkamp, Berlin

Martini M (2017) Algorithmen als Herausforderung für die Rechtsordnung. Juristen Zeitung 72: 1017–1025

Martini M (2018) Artikel 22 DSGVO. In: Paal B, Pauly D (eds) Datenschutz-Grundverordnung Bundesdatenschutzgesetz, 2nd edn. C. H. Beck, München

Martini M, Nink D (2017) Wenn Maschinen entscheiden… — vollautomatisierte Verwaltungsverfahren und der Persönlichkeitsschutz. Neue Zeitschrift für Verwaltungsrecht Extra 36: 1–14

Mayer-Schönberger V, Cukier K (2013) Big data. Houghton Mifflin Harcourt, Boston Merton R (1968) Social theory and social structure. Macmillan, New York

Mittelstadt B, Allo P, Taddeo M, Wachter S, Floridi L (2016) The ethics of algorithms. Big Data Soc 3 (2): 1–21

Montavon G, Samek W, Müller K (2018) Methods for interpreting and understanding deep neural networks. Digital Signal Process 73: 1–15

National Science and Technology Council Committee on Technology (2016) Preparing for the future of artificial intelligence. obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

Neyland D (2016) Bearing accountable witness to the ethical algorithmic system. Sci Technol Hum Values 41 (1): 50–76

OECD Global Science Forum (2016) Research ethics and new forms of data for social and economic research. www.oecd.org/sti/inno/globalscienceforumreports.htm

Palacio S, Folz J, Hees J, Raue F, Borth D, Dengel A (2018) What do deep networks like to see? arxiv.org/abs/1803.08337

Pasquale F (2015) The Black Box Society: the secret algorithms that control money and information. Harvard University Press, Cambridge

Reisman D, Schultz J, Crawford K, Whittaker M (2018) Algorithmic impact assessments: a practical framework for public agency

accountability.ainowinstitute.org/aiareport2018.pdf

Ribeiro M, Singh S, Guestrin C (2016) "Why Should I Trust You?" Explaining the predictions of any classifier. arxiv.org/pdf/1602.04938.pdf

Roth A (2017) Machine testimony. Yale Law J 126: 1972–2259

Rundfunkkommission der Länder (2018) Diskussionsentwurf zu den Bereichen Rundfunkbegriff, Plattformregulierung und Intermediäre.
www.rlp.de/fileadmin/rlp-stk/pdf-Dateien/Medienpolitik/04_MStV_Online_2018_Fristverlaenge rung.pdf

Russell S, Dewey S, Tegmark M (2015) Research priorities for robust and beneficial artificial intelligence. arxiv.org/abs/1602.03506

Sachverständigenrat für Verbraucherfragen (2018) Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren. Gutachten der Fachgruppe Rechtsinformatik der Gesellschaft für Informatik e.V.
http://www.svr-verbraucherfragen.de/wp-content/uploads/GI_Studie_Algorithmenregulierung.pdf

Salmon W (1994) Causality without counterfactuals. Philos Sci 61: 297–312

Sandvig C, Hamilton K, Karahalios K, Langbort C (2014) Auditing algorithms: research methods for detecting discrimination on internet platforms.
www.personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20%2D%2D%20Sandvig%20%2D%2D%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf

Saurer J (2009) Die Begründung im deutschen, europäischen und US-amerikanischen Verwaltungsverfahrensrecht. Verwaltungsarchiv 100: 364–388

Scheppele K (1988) Legal secrets. University of Chicago Press, Chicago

Scherer M (2016) Regulating artificial intelligence systems. Harv J Law Technol 29: 353–400

Scherzberg A (2000) Die Öffentlichkeit der Verwaltung. Nomos, Baden-Baden

Scherzberg A (2013) Öffentlichkeitskontrolle. In: Hoffmann-Riem W, Schmidt-Aßmann E, Voßkuhle A (eds) Grundlagen des Verwaltungsrechts, vol 3, 2nd edn. C. H. Beck, München, § 49

Schwartz B (2015) Google: we make thousands of updates to search algorithms each year.
www.seroundtable.com/google-updates-thousands-20403.html

Selbst A, Barocas S (2018) The intuitive appeal of explainable machines. Fordham Law Rev 87: 1085–1139

Singapore Personal Data Protection Commission (2018) Discussion paper on artificial intelligence and personal data.
www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD%2D%2D-050618.pdf

Stelkens U (2018) § 39 VwVfG. In: Stelkens P, Bonk H, Sachs M (eds) Verwaltungsverfahrensgesetz, 9th edn. C. H. Beck, München

Tene O, Polonetsky J (2013) Big data for all: privacy and user control in the age of analytics. Northwest J Technol Intellect Prop 11: 239–273

Tsoukas H (1997) The tyranny of light. The temptations and paradoxes of the information society. Futures 29: 827–843

Tutt A (2017) An FDA for algorithms. Adm Law Rev 69: 83–123

van Otterlo M (2013) A machine learning view on profiling. In: Hildebrandt M, de Vries K (eds) Privacy, due process and the computational turn. Routledge, Abingdon-on-Thames, pp. 41–64

Villani C (2018) For a meaningful artificial intelligence — towards a French and European Strategy.
www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

von Lewinski K (2014) Überwachung, Datenschutz und die Zukunft des Informationsrechts. In: Telemedicus (ed) Überwachung und Recht. epubli GmbH, Berlin, pp. 1–30

von Lewinski K (2018) Artikel 22 DSGVO. In: Wolff H, Brink S (eds) Beck 'scher Online-Kommentar Datenschutzrecht. C. H. Beck, München

Wachter S, Mittelstadt B, Floridi L (2017) Why a right to explanation of automated decision-making does not exist in the general data

protection regulation. Int Data Priv Law 7: 76–99

Wachter S, Mittelstadt B, Russell C (2018) Counterfactual explanations without opening the Black Box: automated decisions and the GDPR. Harv J Law Technol 31: 841–887

Wexler R (2018) Life, liberty, and trade secrets: intellectual property in the criminal justice system. Stanf Law Rev 70: 1343–1429

Wischmeyer T(2015) Der 》Wille des Gesetzgebers《. Zur Rolle der Gesetzesmaterialien in der Rechtsanwendung. JuristenZeitung 70: 957–966

Wischmeyer T (2018a) Regulierung intelligenter Systeme. Archiv des öffentlichen Rechts 143: 1–66

Wischmeyer T (2018b) Formen und Funktionen des exekutiven Geheimnisschutzes. Die Verwaltung 51: 393–426

Woodward J (2017) Scientific explanation. In: Zalta E (ed) The Stanford encyclopedia of philosophy. Stanford University, Stanford. plato.stanford.edu/archives/fall2017/entries/scientificexplanation

Yudkowsky E (2008) Artificial intelligence as a positive and negative factor in global risk. In: Bostrom N, Ćirkovic M (eds) Global catastrophic risks. Oxford University Press, New York, pp 308–345

Zarsky T (2013) Transparent Predictions. Univ Ill Law Rev 4: 1503–1570

Zweig K (2016) 2. Arbeitspapier: Überprüfbarkeit von Algorithmen. algorithmwatch.org/de/zweites-arbeitspapier-ueberpruefbarkeit-algorithmen

Zweig K (2019) Algorithmische Entscheidungen: Transparenz und Kontrolle, Analysen & Argumente, Digitale Gesellschaft, Januar 2019. https://www.kas.de/c/document_library/get_file?uuid=533ef913-e567-987d-54c3-1906395cdb81&groupId=252038

訳者解説

近年，公的領域か民間セクターかを問わず，様々な領域で意思決定のために AI が活用されるようになっている。例として，金融分野における融資の可否審査や，保険金の支払可否，医療分野における画像診断，雇用分野における人材マッチング，警察分野における顔認証・人物特定などが挙げられよう。

決定結果に疑問や不満がある場合，もし判断を下した主体が人間であれば，その過程および結果について説明を求めることは通常，容易である。しかし，AI の場合にはそれが困難であることが少なくない（いわゆる「ブラックボックス」問題）。ところが，自身について下される判断 ── しかも，金融・医療・雇用などの重大な判断 ── の過程ないし理由を知りえないのであれば，およそ人は自律的な生を送ることが不可能となるのではないか。というのも，通常，人間は自分の行動が将来どのような結果を引き起こすかを知ったうえで，そのメリット・デメリットを自分なりに考慮しつつ，その時々でいかに振る舞うべきかを考える。それが自己決定的・自律的な生の基本である。もし将来，自分が当初予測していなかった不利益を被ることがあれば（例えば，当然受けられるはずの融資を拒否された場合），その理由を知りたがるのは当然であるし，個人にはその権利がある。

以上の背景から，組織の意思決定における AI 利用は個人の自律・主体性を脅かすのではないか懸念されるようになり，実効的な規制をかけようと世界各国で議論が積み重ねられてきている。そして，周知の通り，AI の具体的な規制においては，EU およびその主たる加盟国であるフランス・ドイツが世界を先導している状況にある。

ここに訳出したのは，後述の通り，2020 年にドイツで出版されたハンドブックの一章であるが，この論文の執筆・出版のあとにも AI 規制をめぐって EU では重大な動きがあった。本稿の内容にも関わるので，ここでも簡単に触れておくことにしよう。すなわち，2021 年 4 月になって，EU コミッションが AI の利用を制限する包括的な規制案を公表したのである。この規制案は，AI がもたらすリスクを，①許容不可（禁止），②高リスク，③限定的なリスク，④最小限のリスクの 4 段階に分類し，それに応じて規制をかけることを想定している。とりわけ，②の高リスクのなかには，重要インフラや教育，雇用，クレジット・スコアリング，法執行などの領域で使用される AI が該当するところ，そこでは，開発されたシステムを第三者機関が事前に審査し，登録をすることが予定されている。その際には当然，AI のブラックボックス問題が障壁となる。そのためか，規制案に対してはすでに一部の産業界から批判の声も上がっているようである。いずれにせよ，AI の透明性規制はどの程度，また，いかなる方法により可能となるのか，そして，規制にあたって考慮すべき事項は何か，法学者は真剣に議論しなければならないフェーズに来ていることは間違いない。なお，紙幅の都合から詳述は避けるが，ここで紹介した EU の規制以外にも，各加盟国の国内において規制が検討されているほか，IT 企業の内部や業界団体における自主規制が進められており，アメリカでも，連邦取引委員会（FTC）が，公正さに問題のある AI の取締まりに力を入れ始めているとされる。

このような状況で，まさに AI のブラックボックス問題を正面から取り扱った論文の日本語訳を出版することができたことは誠に幸いである。本論文は，新進気鋭のドイツの公法学者であるトーマス・ヴィッシュマイヤー氏（Thomas Wischmeyer）とティモ・ラーデマッハー氏（Timo Rademacher）が編者となって 2020 年にシュプリンガー社から出版されたハンドブック（『人工知能を規制する（Regulating Artificial Intelligence）』）の一章として書かれたものである。同書は，ドイツの若手の公法研究者が集まって，AI 規制にまつわる広範な論点について手堅い検討を加えたものであり，この分野の研究をする我が国の法学者にとって必読文献といってよいだろう。本論文のなかでも参照されている同書所収の他の論文には「AI とデータ保護基本権」，

「AI と差別」,「AI とソーシャルメディア」,「AI と法執行」などがあり，いずれも興味深いものであるが，上述のような問題意識から，とりわけ重要と考えられた本章を翻訳することを思い立ち，執筆者であり本書の編者でもあるヴィッシュマイヤー氏に連絡をとったところ，翻訳について快諾頂けた。

本稿の原著者について，その経歴と業績を簡単に紹介しよう。ヴィッシュマイヤー氏は 1983 年生まれ，フライブルク大学で法学を学び，州内最優秀の成績で司法一次試験に合格し，2 年間の司法修習を経て 2010 年には司法二次試験に合格，法曹資格を得た。同年よりフライブルク大学のアンドレアス・フォスクーレ（Andreas Voßkuhle）教授（当時，連邦憲法裁判所長官を兼任）の講座で助手を務め，2014 年には『立憲国家の法における目的』と題した論文で同大学より博士号を授与される。2017 年にはビーレフェルト大学のジュニア・プロフェッサー（Junior-professor）に就任し，さらに 2020 年の更新によって同大学正教授に着任している。過去にはイェール大学，ニューヨーク大学でそれぞれ研究・教育に携わった経験も有する。主な専門領域は，憲法・行政法・情報法である。

本稿が扱うテーマは困難であるが，にもかかわらず著者は平易な言葉を用いて理路整然と議論を展開しており，読者が躓く点もそれほど多くないと思われる。AI の利活用に関しては，賛成・反対の両極の立場から熱を帯びた議論が展開されることが多いが，本稿は，AI のブラックボックス性を批判する立場が必ずしも的を射ていないことを指摘し，AI 透明性を実現するための現実的な規制ないし制度設計のあり方を提言している点に，最大の特徴があるといえよう。

翻訳にあたってはまず，1〜2 を栗島が翻訳し，3〜4 の翻訳を小西葉子先生にお願いした。その後，栗島が全体的な訳語の統一等の作業を行った。小西先生は，公私ともにご多忙のなかで無理な依頼を快諾してくださったうえ，翻訳にあたって有益な助言を賜ったことについて，この場を借りて感謝申し上げる。

栗島　智明

※小西担当箇所（3〜4）は，JST SICORP JPMJSC 2107 の支援を受けた成果の一部である。