

ルールと技術の統合的運用としての「法」 LE4SDS（Legal Engineering for Software-Defined Society）の構築に向けて

2026年3月

京都大学大学院法学研究科法政策共同研究センター

* 本報告書は、独立行政法人情報処理推進機構（IPA）及び京都大学大学院法学研究科法政策共同研究センター（KILAP）による、「AI時代のルール（法・標準）とソフトウェアエンジニアリングに関する共同調査研究」の成果の一部を取りまとめたものである。

1. 本研究の目的と背景

1.1 本研究の目的

本研究の目的は、**Software-Defined Society**における新たな法システム（**LE4SDS: Legal Engineering for Software-Defined Society**）の全体構造と、そこでの具体的なテクノロジー活用の在り方（**LT2.0: Legal Tech 2.0**）について、国内外の状況を調査すると共に、技術的及び社会的なフィージビリティも踏まえながら、今後の我が国がとるべき戦略を提言することである。

1.2 本研究の背景

1.2.1 SDSとは

現代は「**VUCA**の時代」とされ、予測困難な状況に迅速かつ柔軟に対応するための「**アジリティ**」の重要性が、ますます高まっている。このアジリティを支える基盤となるのが「**Software-Defined**」の思想である。ソフトウェアを継続的に更新し、不確実性や変化するニーズに柔軟かつ俊敏に対応する考え方であり、さまざまな分野で注目を集めている。これを社会全体に適用したものが「**Software-Defined Society**」であり、**Software-Defined Society**の時代が到来すると見られている¹。

1.2.2 SDSにおける法システムの課題

SDSにおいては、法システムの在り方も変容が迫られる。従来の法律は、リスクが安定した社会において、それらを人間が（十分な注意を払えば）完全にコントロ

¹ ソフトウェアモダナイゼーション委員会 報告書「ソフトウェアのネクストステージに向けて～世界で輝く豊かな日本社会を目指して～」（独立行政法人情報処理推進機構（IPA）、2025年）
<https://www.ipa.go.jp/disc/committee/eid2e00000036bq-att/software-modernization-comittee-20250331-report.pdf>

ールできることを前提にして、そうした人間への行為規範という形で設計されてきた。しかし、そのような既存の法システムには以下のような課題が存在する。

第一に、具体的な行為規範を事前に規定することが困難である。SDSにおいては、技術進歩のスピードが著しく速く、それに伴いリスクの性質や発生態様も急速に変化する。加えて、現代の技術システムは **System of Systems**（複数のシステムが相互接続・相互依存する複合体）としての性格を持ち、個々の構成要素の挙動のみならず、それらの相互作用から生じる創発的リスクをも考慮しなければならない。このような複雑性のもとでは、あらゆる状況に対応した具体的な行為規範を事前に網羅的に規定することは現実的ではない。

第二に、継続的なリスク評価の必要性に既存の法システムが対応できていないことである。従来規制は、製品やサービスの市場投入前の適合性審査（事前規制）を中心としてきた。しかし、SDSにおいては、技術や運用環境が継続的に変化するため、ある時点での適合性評価は急速に陳腐化する。ソフトウェアのアップデート、学習データの追加、利用環境の変化等により、システムのリスクプロファイルは動的に変動する。したがって、一時点での静的な評価ではなく、ライフサイクル全体を通じた継続的なリスク評価が不可欠となるが、既存の法システムはこのような動的評価を想定した設計になっていない。

第三に、安全性・適合性の評価自体が困難である。SDSにおけるリスクは多様な性質を持つ（確率論的リスク、システムリスク、長期的・累積的リスク等）。また、AIシステムに代表されるように、その挙動がブラックボックス化している場合には、なぜ特定の出力が生成されたかを事後的に説明することすら容易ではない。さらに、リスクの測定・評価手法自体が確立途上にあり、何をもって「安全」とみなすかについての社会的合意も形成過程にある。

以上の課題を踏まえると、従来型の規制アプローチ、すなわち、規制当局が事前にコンプライアンスの具体的方法を定義する手法（①）や、事業者がある時点で一定の規範に従うことで適法性を確保する（②）、という手法には限界がある。それでは、コンプライアンスの方法ではなく達成すべき結果（**Outcome**）によって規律できるかという点、その結果自体を明確に定義することが困難であるという根本的な問題がある（③）。「安全なAI」「信頼できるシステム」といった抽象的な目標を、検証可能な形で具体化することは容易ではないのである。

以上の限界を克服する法システムの設計にあたっては、そもそも法が保護しようとしている利益（法益）が何であるかを出発点とし、その法益保護を実現するための手段として、ルールと技術の双方を視野に入れた制度設計が必要となる。第2章では、この「法益保護」を起点とする法システムの枠組み、すなわち **LE4SDS** の基本的な考え方を提示する。

2. LE4SDS の基本的な考え方

LE4SDSとは、データやAIなどのソフトウェアを活用することで、より正確なリスク評価やリスクマネジメントが可能となること、及び法執行機能の自動化による法執行の効率化が可能になることを踏まえた、次世代型の法システム（規制・ソフトロー・責任制裁メカニズム等の総体）を意味する。そこでは、ある法の「法益保護」機能（身体・生命・財産の完全性、プライバシー、信頼できるサイバー空間の保護等）を達成するために、「技術」及び「ルール」が最適な形で統合的に運用され、しかもそれらが、最先端の技術動向を踏まえてタイムリーに更新される。なお、そのようなLE4SDSで活用される技術（PETs、テキスト解析、リアルタイムモニタリング等）を、以下ではLT2.0（Legal Tech 2.0）と呼ぶ。

ここで重要なのは、LE4SDSにおいては、従来のリーガルテックとは異なる思考順序が採られるという点である。従来のリーガルテック（LT1.0）は、「法益保護→人が遵守するべきルール→そのルールを人が履行するための補助技術」という順序で得られてきた。すなわち、法律業務を効率的に遂行するための技術的手法である。これに対して、LE4SDSにおいて活用される技術は、「法益保護→ルールが法益保護との関係で果たしている機能の特定→当該ルールの機能を補完・代替できる技術の特定とそれらを活用するためのルール及び運用方法の特定」という思考順序によって得られる。まさにこの「技術とルールとを法益保護を実現する上で対等な関係とみなす」アプローチが、**Software-Defined Society**の考え方と一致する。

以下、本稿では、LE4SDSのプロセスの仮説を、以下のように整理する。なお、「ルール」とは基本的に法律を想定するが、ルールとしての「法律」と、社会において一定の法益を達成するための社会システムとしての「法」が混同しないようにするために、以下では法律その他の人間向けの規範のことを「ルール」と呼ぶ。

① ルールが果たしている機能の特定

※現在の法規範が法益保護を実現する上で果たしている機能の明確化。

② ルールの機能を代替・補完できる技術（LT2.0）の特定

③ 特定された技術の活用を促すためのルール及び運用方法の設計

たとえば、プライバシー分野においては、以下のような考え方をとる。

<プライバシー分野におけるLE4SDS及びLT2.0の例>

1. 従来の考え方

法の目的（法益保護）：本人の与り知らないところで、第三者（個人情報取扱事業者）が、私的利益の実現のために本人に不利益を与える形で、個人情報を取り扱うことを防ぐことにより、個人情報に関わる本人の法益を保護する。

↓

ルール：個人情報保護法上、原則として、個人情報取扱事業者は本人の同意なく第三者に個人データを提供してはならない。但し、特定の個人を識別できないよう適切に加工した匿名加工情報であれば、本人同意なく第三者に提供できる。

↓

ルールに対応した技術：匿名化処理を行うことで、個人情報を匿名化したまま、より安全にデータを利活用できる。（LT1.0）

↓

しかし、そもそも匿名化したデータでは有用性が不十分である

2. LE4SDS の考え方

法の目的（法益保護）：本人の与り知らないところで、第三者（個人情報取扱事業者）が、私的利益の実現のために本人に不利益を与える形で、個人情報を取り扱うことを防ぐことにより、個人情報に関わる本人の法益を保護する（①と同じ）

↓

ルールが果たしている機能の特定：個人情報保護法上のルールは、本人の関与により、不利益を被るリスクを軽減すると共に、本人に不利益を与えることができない形での提供のみを許可することで、エージェンシー問題（本人に不利益な形で個人情報が使われてしまう問題）を解決しようとしている。

↓

ルールの機能を代替・補完できる技術の特定：個人情報の利活用のライフサイクルにおいて、アクセス管理・暗号化などの **PETs** を組み合わせることにより、エージェンシー問題を解決できるのであれば、本人同意なく提供できる条件を匿名化のみにこだわる必要はないことになる（LT2.0）。

↓

そのような技術の活用を促すためのルール及び運用方法の設計

上記の例において、アクセス管理によりデータ保護を試みる場合には、不正なアクセスを行う者に対し、組織的に責任を持って懲戒等の適切な措置を行うことが必要になる。また、**PETs** を活用する場合には、どのような場面でどのような処理を行うかを文書化して評価すると共に、必要に応じて開示することが重要である。このように、代替技術の活用を促すための技術とルールを最適な形で統合的に運用するための方法（情報開示、適合性評価、トレーサビリティ確保、ベストプラクティスの共有、責任・制裁等のインセンティブ設計、データ基盤の整備等）を確立することが必要となる。（LE4SDS）

3. ケーススタディ

以下では、1. インフラの安全規制、2. 自動運転車の規制、及び3. プライバシーに関する規制について、それぞれ世界各国の関連する取組を分析する。

3.1 <事例 1>インフラ点検における画像識別 AI の活用

3.1.1 概要

我が国では、デジタル庁の主導によって、目視点検・定期検査義務・常駐義務などのいわゆる「アナログ規制」の撤廃が進められており、2025年12月末までに

8000を超える規制や規則、見直しが完了している²。これによって、たとえばインフラの点検や事業所のセキュリティ確保などを実施するにあたって、AIを用いたドローンや監視システムを導入することが法的に認められることになった。

従来の法令が、「法目的（施設の安全性やセキュリティ等）を達成するために必要なアナログ的手法をルールによって特定する」というアプローチを採っていたのに対し、改正後の法令は、「法目的を達成するための技術を事業者が自由に選定してよい」というアプローチを採っている。また、その規制目的の達成に有用なテクノロジーマップも公表されている³。これは、上記のLE4SDSの考え方のうち、①ルールが果たしている機能の特定、及び②ルールの機能を代替・補完できる技術（LT2.0）の特定が一部遂行されている状態といえる。

しかし他方で、本改正では、特定の技術の導入に伴って、どのような安全を達成すべきかということや、そのためにどのような技術評価・組織体制・情報開示を行うべきかといったことについては、原則として明確化されていない。そして、この点が明確にならなければ、事業者にとっては技術活用に伴う法的リスクが不明確となり、積極的に技術活用することは難しい。

この課題について、シンガポールにおける建築物等のインフラ点検に関するルールが参考になる。

3.1.2 シンガポール：建築物ファサード点検におけるドローン活用

3.1.2.1 従来のルールが果たす機能とその課題

シンガポールでは、建築物のファサード（外壁）からの落下物による人身事故を防止するため、築20年以上の建築物に対して7年ごとの定期点検が義務付けられている⁴。

こうした規制は、有資格の点検員による目視・打診点検を通じて、タイルの浮き、ひび割れ、構造的劣化等の欠陥を検出し、落下事故を未然に防止するという法益を実現するためのものである。

しかし、従来の方法には重大な課題があった。高層建築物の点検には足場の設置やゴンドラの使用が必要となり、1棟あたり数万～数十万シンガポールドルのコストと数週間の作業期間を要していた。また、高所作業に伴う労働災害リスクも無視できない。

² デジタル庁「アナログ規制の見直しに係るフォローアップ実績」。

<https://www.digital.go.jp/policies/digital-extraordinary-administrative-research-committee#follow-up-results>

³ デジタル庁「テクノロジーマップ」。<https://www.digital.go.jp/experimental/technology-map/>

⁴ Building Control Act (Cap. 29, 2020 Rev Ed) 及び Building Control (Periodic Inspection of Buildings) Regulations 2021。シンガポール建築建設庁（BCA）は2022年1月1日より建築物ファサードの定期点検制度（Periodic Facade Inspection: PFI）を施行した。築20年以上の建築物に対して7年ごとの点検が義務付けられている。

3.1.2.2 ルールの機能を代替・補完できる技術（LT2.0）の特定

このような課題を解決するためには、人間による目視点検ではなく、ドローンとカメラを用いた点検を可能とすることが考えられる。そこで、シンガポール政府は、上記の法が果たした機能を維持しつつ従来手法の課題を解決する技術として、一定の条件の下で、ドローン撮影画像に対するAIによる欠陥検出が可能になるようにした。その際の参照規格として、TR 78 シリーズ（Technical Reference 78）という規格が整備された。

- TR 78-1:2020：ドローンを用いたファサード点検のための要件
（Requirements for façade inspection using unmanned aircraft systems）⁵

本規格は、ドローン（UAS）を用いたファサード点検のプロセス全体を規律する。点検準備段階（リスクアセスメントの実施、飛行許可の取得、保険加入）、ステークホルダーからの承認取得（飛行計画書・点検計画書の提出）、実施段階（飛行の安全管理）、事後処理（データ処理・報告書作成）という一連のフェーズを規定する。そして、各フェーズにおいて、撮影画像の品質基準（たとえば、可視光カメラは0.15 cm/pixel以下の解像度を推奨）が定められている。また、ドローンによる撮影は不可避免的に周辺の人物や車両を撮影しうることから、個人情報保護（プライバシー保護のための撮影データ管理・匿名化措置）に関する詳細な要件も設けられている。サーモグラフィやLiDAR（レーザー光を用いた空間把握システム）等の高度センサーを使用する場合には、適格な専門家を別途関与させることを求めており、技術の適正利用を担保している。有資格の点検員（Competent Person: CP）は引き続き、点検の監督、調査結果の評価、報告書への署名・承認に責任を負う。

- TR 78-2:2021：品質管理およびAI適用に関する仕様（Specification for quality management and application of artificial intelligence (AI)）⁶

本規格は、上述のTR 78-1:2020を前提として、画像識別AIを活用した欠陥検出の品質管理体制と評価基準を規定する。特に重要なのは、AI評価（アセスメント）制度の導入である。AIシステムは、規制当局またはその指定機関によって任命されたAI評価機関（AI assessor）による事前評価に合格しなければならない。評価基準として、コンクリート・左官材のひび割れ、剥落、金属腐食等の欠陥を少なくとも各60%以上、全体平均で75%以上の再現率（recall ratio）で検出できること、また、プライバシーについては、個人（顔・容貌）や車両の匿名化処理を100%実施できることが要求される。AIシステムを改変した場合には再評価が必要となり、継

⁵ Enterprise Singapore, TR 78-1:2020+C1:2021, “Building facade inspection using unmanned aircraft systems (UAS) — Part 1: Requirements for facade inspection using unmanned aircraft systems,” Singapore Standardisation Programme.
<https://www.singaporestandardseshop.sg/Product/SSPdtDetail/cd12a4dd-10db-4f34-b1b2-e7ed574393ba>

⁶ Enterprise Singapore, TR 78-2:2021, “Building facade inspection using unmanned aircraft systems (UAS) — Part 2: Specification for quality management and application of artificial intelligence (AI),” Singapore Standardisation Programme.
<https://www.singaporestandardseshop.sg/Product/SSPdtDetail/27c00227-b76d-426b-bfc0-b503eff6bc54>

継続的な品質維持を確保する仕組みが構築されている。また、AIシステムを用いる点検サービス事業者（Inspection Service Provider: ISP）には、技術管理者の資格要件（ファサード点検資格証明書の取得、ドローンインフラ点検の実務経験2年以上）が課されており、専門的責任体制が維持されている。TR 78-2が扱う画像識別AIは、従来のルールの機能を以下のように技術的に代替・補完する。

規制が求める機能	技術による代替・補完（TR 78-2）
目視による欠陥検出	AIによる自動欠陥検出 （ひび割れ、剥落、金属腐食等）
個人情報・プライバシー保護	AIによる個人・車両の自動匿名化

（出典：TR 78-2:2021 Table B.1）

3.1.2.3 特定された技術の活用を促すためのルール及び運用方法の設計

このようなドローン・AI技術は、「Competent Personによる包括的な目視点検および打診点検（hands-on inspection）」を全面的に代替するものとは位置付けられておらず、あくまで有資格者が選択的に活用できる補助的手段として規格化されている。すなわち、人間によるモニタリング（Human-in-the-loop）を前提とする制度設計が採用されている。AI技術は欠陥候補の検出と優先順位付けに用いられるが、最終的な判定と報告書への署名は依然として有資格者（CP）が行い、最終的な責任を負うのも人間である。

その上で、検査に用いるAIについては「承認制（事前評価・承認制度）」が採用されている。すなわち、TR 78-2:2021は、AIシステムを点検業務に使用するためにあたり、欠陥検出性能（各60%以上かつ全体平均75%以上）と匿名化処理性能（100%）の双方について、規制当局またはその指定機関による事前評価への合格を義務付けている。この評価は定期的な再審査（2年以内ごとの再評価が推奨）を通じて継続的に実施される。

3.1.3 性能基準及び評価プロセスを明確化することのメリットと課題

シンガポールのドローン撮影及び画像識別AIを用いたファサード点検に関するルールをまとめると、以下のようになる。

- ドローン・AIは人間の作業を代替するものではなく、人間による検査を補完するものと位置づけられる。
- ドローン・AIを用いた検査を実施するための資格要件が定められている（ファサード点検資格証明書の取得、ドローンインフラ点検の実務経験2年以上等）
- 検査に用いるAIに対しては、規制当局または指定機関による事前評価及び承認制が採られている。AIの満たすべき性能については、規格において、欠陥の検出率や人物画像等の匿名化等の観点から基準が設けられている。

このようなルールは、我が国のアナログ規制撤廃後の制度設計に対して、以下の示唆を与える。

第一に、ファサードの安全性の確保という法益を実現するために活用できる技術について、規格を通じて明確な性能を定義している点である。すなわち、ドローンやAIをインフラ点検等に活用する場合、単に「使用を許容する」ととどまらず、欠陥種別ごとの検出再現率（各60%以上かつ全体平均75%以上）、個人・車両の匿名化処理の達成率（100%）、撮影画像の最低分解能（GSD 0.15 cm/pixel以下）といった定量的な性能基準を設けているのである。他方、我が国の「アナログ規制撤廃」後の規制のほとんどは、コンプライアンス目的で使われる技術の性能の要求を明確にしておらず、具体的にどのような性能のAIシステムを導入するかは事業者の判断に委ねられている。このようなアプローチは、自身で適切なリスク判断ができる事業者にとっては柔軟で使いやすい制度である一方、AI技術の性能評価に関する専門的知見に限られた事業者にとっては、新技術を導入する際の障害ともなり得る。そのため、シンガポールのように一定の性能基準を明確にするアプローチは、我が国でも検討に値する。

第二に、AIの事前承認・定期的再評価制度を導入していることである。シンガポールは、AIシステムを規制当局指定機関による定量的評価（再現率基準）に合格させた上で初めて法的点検業務への使用を認める制度を設けている。さらに、TR 78-2:2021の附属書Bにおいては、評価用テスト画像の形式（JPEG（品質95%以上）またはPNG）・分解能（GSD 0.15 cm/pixel以下）、AI能力ごとのサンプル画像枚数（25～100枚）、AIシステムの処理時間制限（2時間以内）といった具体的な評価プロセスが詳細に規定されており、客観性と再現性の高い評価が担保されている。また、最後の評価から2年以内ごとの再評価実施が推奨されている。我が国でも、使用されるAIの評価プロセスをルールとして明確化することが、技術代替の社会的信頼性を確保する上で重要となる。

他方で、シンガポールのアプローチには限界もある。

第一に、この制度が、あくまでもAIを補助的に導入するのにとどまり（Human-in-the-loop）、最終的な責任は依然として人間に留保されたままという点である。このような対応は、急速な技術進化に伴うリスクに対する合理的なヘッジであり、ハイリスク分野における過渡期的な対応としては適切なものといえる。しかし、後述の自動運転技術のように、人間の有資格者よりもAIの方が高いパフォーマンスを発揮できる領域（建物の欠陥の検出もAIの方が優れている可能性は高い）においては、個々の検査者に責任を負わせることは却って安全性を低下させることになりかねない。そのため、AIシステムが点検業務において有資格者を上回る性能を客観的に実証できる段階に至った場合には、現行のHuman-in-the-loopを前提とした責任の在り方についても、制度的な再設計を検討することが、真の安全性向上に資するものと考えられる。

第二に、このような定量的な性能基準や客観的な評価プロセスが可能となるのは、ファサード点検のように、①検出すべき欠陥の種別があらかじめ列挙可能であ

り、②入力データの形式（静止画像）と品質要件（分解能）が規格化でき、③合否判定の基準（再現率）を数値で表現できる、という三つの条件が揃う業務に限られる。自動運転や生成AIのように、入力空間・リスクシナリオが事実上無限であり、「何をもって合格とするか」を事前に網羅的に定義することが原理的に困難な領域では、同様のアプローチをそのまま適用することは難しい。この意味で、シンガポールのTR 78シリーズが採用した「性能基準の事前規格化+承認制」は、業務スコープの限定性を前提とした制度設計であり、その射程を見極めた上で我が国への応用可能性を検討することが求められる。

3.2 <事例2>自動運転車に関するルールの設計

3.2.1 概要

事例1のインフラ点検が「人の代わりにシステムを活用する」という文脈での技術代替であったとすれば、自動運転車（Automated Vehicle: AV）の問題は、これを一段と深化させたものと理解できる。なぜなら、インフラ点検においては依然として人間（有資格の点検員）が最終判断者として残存していたのに対し、SAEレベル4以上の自動運転⁷においては「運転操作の主体としての人間」がもはや存在しないためである。

従来の道路交通法制は、「人が運転する」ことを前提として設計されており、安全確保・法的責任・法執行のすべての機能が「運転者」という人間的行為者の存在に依拠している。自動運転技術の高度化に伴い、この前提が根底から崩れることとなる。この事例では、かかる状況において従来のルールが果たしてきた機能を特定した上で、それを代替・補完する技術とその活用を促すためのルールの設計について、英国のAutomated Vehicles Act 2024（以下「AVA」）を主要な参照対象として分析する。

3.2.2 従来のルールが果たす機能とその課題

道路交通法制が果たしてきた機能は、以下のように整理できる。

- (a) 安全確保機能：運転免許制度・保安基準・交通規則（信号遵守、速度制限、優先権ルール等）を通じて、道路上の安全を確保する。これは、運転者が一定の知識・技能・注意力を有することを前提とし、その行動を規範によって規律することで実現される。
- (b) 法的責任の帰属機能：交通事故が発生した場合、運転者の過失の有無を評価し、損害賠償・刑事責任の帰属先を決定する。この機能は、「意思をもった人間の行為」に対して過失を問うという法システムの根本的構造に基

⁷ SAE International, J3016_202104, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," April 2021.

づいている。すなわち、「制裁を受けることを恐れ、行動を改める意思ある人間」の存在が前提となっている。

しかし、自動運転技術の高度化はこれらの前提を揺るがす。第一に、安全確保の機能についていえば、AIによる状況判断・操作と人間のそれとは根本的に異なるため、何をもって安全と評価するかという基準や評価手法が、従来型の免許・保安基準の体系の中には存在しない。第二に、法的責任の帰属については、自動運転システムが人間の個別の「意思」に基づかない形で挙動を決定する以上、伝統的な過失責任の概念を適用することが困難になる。「誰が何を誤ったか」という問いに対して、AIシステムの特性上、明確に答えることができない場合が生じる。さらに、交通違反の主体がもはや意思ある人間ではなく自律的なシステムである以上、違反への制裁が本来の安全行動への誘因として機能しない。

加えて、自動運転システムの安全性は導入時点での固定的な性質ではなく、ソフトウェアの継続的な更新・学習データの蓄積・稼働環境の変化によって動的に変動する。したがって、事前の型式認証のみで安全性を担保することには根本的な限界がある。

3.2.3 ルールの機能を代替・補完できる技術（LT2.0）の特定

上記の課題に対応するために活用が期待される技術は、自動運転システムそのものだけでなく、その安全性を継続的に監視・評価するための一連の技術群である。具体的には、（i）走行データの継続的収集・分析技術（走行ログ、センサーデータ、インシデント記録等のリアルタイム蓄積）、（ii）AIシステムの性能評価・シミュレーション技術（仮想環境でのエッジケーステスト、適合率・再現率等の定量指標による評価）、（iii）異常検知・ハザード報告技術（通常走行中のリスク事象の自動検出とフラグ付け）が挙げられる。

これらの技術は、AIシステムの定量的な性能基準への適合評価と、稼働後の継続的なデータ収集・モニタリングによって、従来人間への免許制度が担っていた安全確保機能を代替する。また、走行ログ・インシデントデータが「何が起きたか」の客観的記録を提供することで、事後的な原因究明と責任評価を可能にする。

ただし、ここで問題となるのは、これらの技術を活用するための規範的な枠組みが自明ではないという点である。すなわち、「どのような基準・プロセスで自動運転システムの安全性を評価・承認するか」「継続的なデータ収集・モニタリングをどの主体がどのような義務として担うか」「事故発生時に開発者・製造者・運行事業者・インフラ管理者のいずれがいかなる根拠で責任を負うか」といった問いに対して、従来の法制度は解をもたないことである。ファサード点検とは異なり、自動運転は入力空間・リスクシナリオが事実上無限であり、性能基準の網羅的な事前規格化には原理的な限界がある。そのため、技術の活用を促すための規範設計においては、事前の型式承認だけではなく、稼働後の継続的モニタリング・ハザード報告・定期的再評価を組み込んだ動的な安全確保の仕組みと、操作する人間の「意

思・過失」に依拠しない新たな責任帰属の枠組みを、一体的に制度化することが不可欠となる。次節では、この課題に対する英国 AVA のアプローチを分析する。

3.2.4 ルール及び運用方法の設計：英国 Automated Vehicles Act 2024

（1）立法の背景と設計思想

英国の AVA（Automated Vehicles Act 2024）は、2024年5月に成立した⁸。2027年の完全施行を目指して現在 CCAV（Centre for Connected and Autonomous Vehicles）が中心となって二次立法の整備が進められており、2026年春には自動旅客サービス（Automated Passenger Services）の先行パイロットも予定されている⁹。

AVA の設計思想上の最大の特徴は、SAE レベル等の技術的分類に依拠せず、「自己走行能力（self-driving capability）」の有無という機能的基準を採用した点にある。すなわち、車両が走行中に人間による継続的な監視・介入を要せず自律的に走行できるか否かを問い、そのような能力を持つと認定された車両のみを「認定自動運転車（Authorised Automated Vehicle）」として規律する枠組みを採っている。

（2）「自己走行テスト」と「安全原則声明（Statement of Safety Principles）」認定を受けるためには、すべての自動運転車が自律走行テスト（self-driving test）に合格しなければならない。このテストの核心は、車両が「安全かつ合法的に」自律走行できることの確認であり、その評価基準として、交通大臣（Secretary of State for Transport）が「安全原則声明（Statement of Safety Principles: SoSP）」を策定し議会に提出することが法律上義務付けられている（AVA 第2条）¹⁰。SoSP は、「認定自動運転車が、慎重かつ有能な人間運転者（careful and competent human driver: C&C human driver）と同等以上の安全水準を達成すること」、かつ「自動運転車の普及によってグレートブリテンの道路安全が向上すること」を確保する観点から策定されなければならない。

この SoSP における「C&C human driver」基準は、事例1で扱ったインフラ画像識別 AI が採用していた定量的アプローチとは根本的に異なり、機械可読な数値基準への変換が極めて困難である。すなわち、画像識別 AI の場合、評価対象となる AI の機能が「ファサードという特定種別の欠陥を静止画像から検出する」という限定されたタスクであるため、入力空間と合否基準を事前に網羅的に定義できた。これに対して自動運転の場合、「慎重かつ有能」という概念はそもそも人間の行動規範として発展してきた法的概念であり、無限に近いリスクシナリオ・走行環境の組み

⁸ Automated Vehicles Act 2024, c. 10. Royal Assent は 2024 年 5 月 20 日。

<https://www.legislation.gov.uk/ukpga/2024/10>

⁹ UK Department for Transport, “Automated Vehicles Act 2024 implementation,” Written Ministerial Statement, 10 June 2025. <https://www.gov.uk/government/speeches/automated-vehicles-act-2024-implementation>

¹⁰ Automated Vehicles Act 2024, s. 2. 同条は、交通大臣が「statement of safety principles」を策定し議会に提出することを義務付ける。

合わせの中で、何をもってその基準への適合とみなすかを事前に数値で定義することは原理的に困難である。

こうした定量化の困難に対して、英国が採用している代替的アプローチは以下の三点に集約される。

第一に、安全ケース（**safety case**）アプローチと走行領域を限定した評価である。**ASDE**は、自社の自動運転システムが展開する特定の走行領域（**Operational Design Domain: ODD**）ごとに、安全性を実証するための独自の方法論・評価指標を開発し、規制当局に対してその妥当性を説明する責任を負う構造を採る。統一された定量的閾値を一律に規定するのではなく、各社が自らの技術・**ODD**特性に応じた評価根拠を提示するという、原則ベースの規制設計である。

第二に、事前評価と事後モニタリングの組み合わせ（ライフサイクル型安全保証）である。**AVA**は事前の型式認証のみに依拠するのではなく、稼働後の継続的な「**in-use assurance**」制度を法律上の柱として位置づけている（**AVA**第38条）¹¹。具体的には、先行指標（**leading metrics**：事故発生前のリスク事象の検出・ハザード報告）と遅行指標（**lagging metrics**：衝突・死傷者数等の事後的結果）の双方を用いた継続的モニタリングを行い、**ASDE**には走行データ・インシデントデータの報告義務が課される。さらに、毎年の全体的な安全パフォーマンス評価（**annual assessment**）が義務付けられており、「一度認定すれば終わり」ではなく全ライフサイクルにわたる継続的な安全確認という設計思想が貫かれている。

第三に、国際的な基準調和への参画である。英国は**UNECE**（国連欧州経済委員会）において自動運転システム（**ADS**）に関する新たな国際型式認証規則の策定を主導しており、この規則は「不合理なリスクからの自由（**free from unreasonable risk**）」という原則ベースの基準を軸に構築される方向にある。特定の数値閾値を一国が独自に設定するのではなく、国際的な基準調和の枠組みの中で安全要件を発展させるというアプローチである。

なお、**SoSP**の具体的内容は2025年6月から9月にかけての意見募集を経てなお策定中であり、詳細なコンサルテーションが2026年夏に予定されている¹²。「**C&C human driver**」基準の定量化・機械可読化は依然として世界的な課題として残されている。

（3）新たな法的主体の設計：ASDE・UiC・NUiC事業者

従来の道路交通法制における責任体系は、「意思をもった人間（運転者）の行為」に対して過失を問うという構造を前提としている。しかし自動運転システムは、開発者・利用者が個別・具体的な挙動を意思的に制御できない形で挙動を決定

¹¹ Automated Vehicles Act 2024, s. 38. 同条は、認定自動運転車の一般的パフォーマンスを監視・評価するための制度（**in-use assurance**）の整備を規定する。

¹² UK Department for Transport, “Automated vehicles: statement of safety principles — Call for Evidence,” 10 June 2025. <https://www.gov.uk/government/calls-for-evidence/automated-vehicles-statement-of-safety-principles>

する。このため、「誰の何の過失か」という問いに原理的に答えられない場面が生じる。AVAはこの問題に対し、自動運転に固有の新たな法的主体を創設することで、「意思ある人間の過失」に依拠した従来型の責任体系を抜本的に再設計している。これは個人の過失への帰責から、組織としての継続的な安全確保義務の履行という問題へと責任構成を転換する発想に基づくものである。

ASDE (Authorised Self-Driving Entity) は、個別の認定自動運転車に対応して指定され、その車両が継続的に自己走行テストを充たすことに対する包括的な責任を負う主体である。車両メーカー、ソフトウェア開発者、またはその合弁体が想定されており、すべての認定 **AV** と **ASDE** の対応関係は公開登録簿に記載される。自動運転機能の作動中に発生した交通違反・事故については、従来の「運転者」に代わり **ASDE** が主たる責任主体となる。また **ASDE** は「指名管理者 (nominated manager)」を選任し、この指名管理者が規制当局への情報提供に係る個人的責任を担う。

UiC (User-in-Charge) は、自動運転機能が作動している間も車内に乗車し、必要な場合に操作を引き継げる位置にいる個人である。自動運転機能作動中の運転行為に起因する違反については原則として免責されるが、車両の整備状況の確保・保険の維持・14歳未満の乗客へのシートベルト着用義務等、「運転者」としての一部の義務は引き続き負う。また、移行要求 (transition demand) が発出された後の移行期間内に事故が発生した場合には、**UiC** への責任帰属も生じ得る。

NUiC 事業者 (Licensed No User-in-Charge Operator) は、車内に操作引継者がいない完全無人走行形態の自動運転車の運行を監督する免許事業者である。**ASDE** (技術・認定の責任者) と **NUiC 事業者** (運行監督の責任者) を分離しつつ連携させるという二層構造は、責任の所在を技術面と運用面で明確に分節化する設計として注目される。

なお、**UiC** から **NUiC** への段階的移行という制度設計は、事例1で提起した「AIが人間を上回る性能を発揮できる段階に至った場合の **Human-in-the-loop** の在り方」という問題に対する一つの制度的応答として理解できる。**AVA** は、**Human-in-the-loop (UiC)** を前提とする段階と、**Human-in-the-loop** を前提としない段階 (**NUiC**) の双方に対応可能な制度的枠組みを用意することで、技術の成熟度に応じた責任設計の段階的移行を可能としているのである。

このように、**AVA** は「誰の意思・行為が事故を引き起こしたか」という過失犯的な問いから離れ、「どの組織がどの安全確保義務を継続的に担うか」という組織的・継続的な責任の枠組みへと移行することで、人間の意思が直接介在しない自動運転という技術の現実に法制度を適合させている。

(4) 情報開示義務と刑事制裁

自動運転システムの安全性は、**ASDE** や事業者からの情報提供に大きく依存せざるを得ない。**AVA** はこの点に関して、制裁とインセンティブを組み合わせた制度設計によって、事業者による誠実な情報開示を担保している。

制裁面では、規制当局が **ASDE** や免許 **NUiC** 事業者に対して情報提供を要求し施設への立入調査を行う権限が付与されている。さらに、虚偽情報の開示・情報の不開示に対する刑事制裁が新設され（最高禁固5年または罰金）、法人だけでなく情報開示に関与する「指名管理者」および「関連上級管理職（**relevant senior manager**）」といった個人も刑事訴追の対象とされる（**AVA** 第24～26条）¹³。すなわち、法人レベルの義務と個人レベルの刑事責任を並列して課すことで、組織内部における情報の隠蔽・過小報告に対して二重の制裁圧力をかける構造となっている。

インセンティブ面では、**デュー・ディリジェンス・ディフェンス（due diligence defence）** が整備されており、企業・個人が義務違反を避けるための合理的な予防措置を適切に実施し相当の注意を払ったことを立証した場合には免責される（**AVA** 第24条第7項・第25条第2項）¹⁴。これにより、安全情報を誠実に開示・共有する事業者が法的保護を受け、隠蔽を図る事業者が法的リスクを負うという非対称な誘因構造が生じる。

（5）マルチステークホルダー協働の制度化：**Zenzic** の役割

AVA の実施プロセスにおいて特筆すべきは、政府・産業界・学術機関が同比率で共同出資して設立された **Zenzic** という第三者機関の存在である¹⁵。**Zenzic** は、自動運転車の安全確保に向けたステークホルダー間の対話・実証を支援する場として機能しており、実証環境（試験場・公道試験）の整備と官民コミュニケーションの促進を担っている。法制度の形成過程において事業者・市民・研究者が継続的に関与できるマルチステークホルダー型の仕組みを制度化したことは、**アジャイルガバナンス** の観点から重要な先例を提供している。

3.2.5 我が国の制度設計への示唆

英国の **AVA** の分析から、我が国の自動運転規制の設計に向けた示唆として以下の点が導かれる。

第一に、**SoSP** アプローチの採用とそれを支える公的テスト環境の整備である。3.2.3 で述べたように、自動運転のリスクシナリオは事実上無限であり、シンガポールの **TR 78-2** が採用した「定量的閾値の事前規格化+承認制」をそのまま適用することには原理的な限界がある。英国が示すように、事前の型式認証と事後モニタリングを組み合わせたライフサイクル型の安全保証を制度化しつつ、**C&C human driver** を **ODD** ごとにシナリオベースで評価するアプローチに頼らざるを得ない。し

¹³ **Automated Vehicles Act 2024, ss. 24-26**. 第24条は虚偽情報の提供・重要情報の不開示に対する犯罪を、第25条は法人犯罪における上級管理職の個人的刑事責任を、第26条は最高刑（禁固5年または罰金、あるいはその併科）を規定する。

¹⁴ **Automated Vehicles Act 2024, s. 24(7) 及び s. 25(2)**. **デュー・ディリジェンス・ディフェンス** の要件として、被告が義務違反を回避するための合理的な予防措置を講じ、相当な注意を払ったことの立証を求める。

¹⁵ <https://zenzic.io/>

しかしこのアプローチが実効的に機能するためには、各 ODD における自動運転システムの安全性を客観的に評価するためのシナリオ設計・データ収集・シミュレーション環境が不可欠である。国土交通省の 2025 年 3 月報告書¹⁶においても、シナリオベース安全性評価手法の導入と C&C human driver の安全要件の具体化の方針が示されており、これを実施するための官民協働の共通テスト環境を国際的な議論と連動しながら整備することが重要である。

第二に、「過失」概念の再構築と責任主体の明確化である。3.2.2 で確認したとおり、自動運転システムは人間の個別の意思に基づかない形で挙動を決定するため、「誰の何の過失か」という従来の過失犯的な問いには原理的に答えられない場合が生じる。我が国においては過失および因果関係が比較的広く認められる傾向にあり、開発者個人への過失犯としての訴追リスクが自動運転開発の萎縮要因となりかねない。英国は ASDE という組織的・継続的な安全確保責任の帰属主体を法的に創設することでこの問題に対処している。我が国においても、開発者・製造者・運行事業者のいずれが何に対して継続的責任を負うかを制度的に明確化し、個人の過失責任への過度な依拠を避ける制度設計への移行が、自動運転の国際競争力確保および社会実装の観点から有益である。

第三に、情報開示インセンティブの制度的設計である。3.2.3 で指摘したように、走行ログ・インシデントデータが事後的な原因究明と責任評価を可能にする前提として、事業者が誠実にデータを開示することが不可欠である。しかし事業者にとっては、インシデント情報の積極的な開示が法的・競争的リスクを高める可能性があり、隠蔽・過小報告への誘因が生じやすい。AVA が採用した、法人と個人（指名管理者・関連上級管理職）双方への刑事制裁とデュー・ディリジェンス・ディフェンスを組み合わせる設計は、誠実に開示する事業者が保護され隠蔽を図る事業者が制裁を受けるという非対称な誘因構造を制度的に形成するものである。我が国においても同様の枠組みの整備が、「安全文化」を官民協働で構築する上で有効な手段となり得る。

第四に、マルチステークホルダー型の常設対話プラットフォームの設立である。自動運転の安全基準が世界的にも未確定である以上、技術・法制度・運用の三者が継続的に対話しながら制度を動的にアップデートしていくことなしには、上記第一～第三の課題を継続的に解決し続けることはできない。Zenzic のように政府・事業者・研究者が同等の立場で継続的に意見交換を行う常設の場の設立が、LE4SDS フレームワークにおける「③技術の活用を促すためのルール及び運用方法の設計」をアジャイルに更新し続けていく上で不可欠の基盤となる。

¹⁶ 国土交通省「自動運転車の安全性評価手法に関する検討会」報告書等（2025年3月）。

3.3 <事例3>個人情報保護制度の法機能分析：同意ベース規律の限界とリスクベース・アプローチへの転換

3.3.1 概要

事例1（インフラ点検）は特定タスクの技術代替に伴う規制設計の課題（点検の手段の問題）を、事例2（自動運転）は自律的システムによる人間の全面的代替に伴う安全確保と責任帰属の再設計の問題を、それぞれ論じてきた。事例3として取り上げる個人情報保護の問題は、これらをさらに本質的な次元に深化させるものである。なぜなら、規制の対象が「物理的な作業の安全性」ではなく「情報の取扱いが人間の権利利益に与える影響」であるため、「何がリスクであり何が許容されるか」という価値判断そのものが技術の発展に伴って動的に変化するからである。すなわち、事例3は特定のタスクの代替ではなく、法益保護の在り方そのものが問われる事例である。

本節では、この事例にLE4SDSの三段階プロセスを以下のように適用する。まず、現行の個人情報保護法の手続的規律が法益保護との関係でいかなる機能を果たしており、そこにいかなる限界があるかを分析する（①ルールが果たしている機能の特定：3.3.2）。次に、その機能を代替・補完しうるプライバシー保護技術（PETs）を特定する（②技術の特定：3.3.3）。最後に、それらの技術の活用を促すためのルール及び運用方法の在り方を、国際的な制度設計も参照しつつ検討する（③ルール及び運用方法の設計：3.3.4）。

3.3.2 現行の個人情報保護法が果たしている機能とその限界

（1）現行法の手続的規律の構造

個人情報の保護に関する法律（平成15年法律第57号。以下「個人情報保護法」という）は、端的にいえば、本人の与り知らないところで、第三者が私的利益の実現のために本人に不利益を与える形で個人情報を取り扱うことを防止することにより、「個人の権利利益を保護すること」¹⁷を目的としている。個人情報保護法は、この目的を達成するために、事業者に対して、本人の関与の機会を確保するという手続的規律を基軸とした設計を採用している。手続的規律の代表例が、（i）個人情報取得時の利用目的の特定・通知等¹⁸、（ii）本人同意のない目的外利用の禁止¹⁹、（iii）第三者提供時の本人同意の取得²⁰等である。このような個人情報保護法の手続的規律には、一定の例外が設けられている。そのうちの1つが、委託先への提供である²¹。委託先への提供は、委託された業務の範囲内でのみ、本人との関係に

¹⁷ 個人情報保護法第1条。

¹⁸ 個人情報保護法第17条、第21条。

¹⁹ 個人情報保護法第18条第1項。

²⁰ 個人情報保護法第27条第1項。

²¹ 個人情報保護法第27条第5項第1号。

において提供元である事業者と一体のものとして取り扱われることに合理性があるため本人同意が不要とされる²²。また、第三者提供規制の対象は「個人データ」²³であるが、個人情報保護法の定めに従い特定の個人を識別できない形に加工した「匿名加工情報」²⁴であれば第三者提供に際し本人同意は不要である。さらに、個人情報保護法の定めに従い他の情報と照合しない限り特定の個人を識別することができないように加工した「仮名加工情報」²⁵は本人同意なく目的外利用できる。これらの手続的規律の例外は、法益侵害リスクが顕在化する可能性が低いと評価される処理を制度的に許容するものともいえるが、後述のとおりいずれも技術的担保措置とは切り離された設計となっている。

こうした手続的規律が法益保護との関係で果たしている機能は、本人を関与させることにより、不利益を被るリスクを軽減すると共に、本人に不利益を与えることができない形での情報処理のみを許容することで、エージェンシー問題（本人に不利益な形で個人情報が使われてしまう問題）を解決することにある。

（2）手続的規律の課題：過剰規制と規制漏れ

しかし、現行の個人情報保護法（以下「現行法」という）の手続的規律は、法益保護という観点から二つの課題に直面している。

第一に、「過剰規制」の側面である。現行法は、個人の権利利益への直接の影響が低いと評価される処理においても、形式的に本人同意を要求する。たとえば、委託先が複数の委託元から提供を受けた個人データを本人ごとに突合（名寄せ）することなく、サンプルとなるデータ数を増やす目的で合わせて1つの統計情報を作成することは、本人同意なく可能であるが、当該個人データを本人ごとに突合（名寄せ）し、本人ごとに個人データの項目を増やす等した上で統計情報を作成する場合、本人同意が必要とされる²⁶。このように、複数の事業者が保有するデータを横断的に集計、分析し、統計情報を作成する場合において、実質的なリスクの有無とは切り離された形で同意を要求することは、データ利活用に対する合理的根拠のない障壁となりうる。

第二に、「規制漏れ」の側面である。本人同意を取得したという事実は、その後の処理により生じ得る法益侵害リスクを規律するものではない。たとえば、AIによる自動的な評価・スコアリング・プロファイリングが生み出すリスク、すなわち本

²² 個人情報保護委員会「個人情報の保護に関する法律についてのガイドライン（通則編）」3-6-3（1）。

²³ 個人情報データベース等を構成する個人情報という（個人情報保護法第16条3項）。「個人情報データベース等」とは、個人情報を含む情報の集合物であって、特定の個人情報を電子計算機を用いて検索することができるように体系的に構成したもの、又は、特定の個人情報を容易に検索することができるように体系的に構成したものとして政令で定めるものをいう（個人情報保護法第16条第1項）をいう。

²⁴ 個人情報保護法第2条第6項。

²⁵ 個人情報保護法第2条第5項。

²⁶ 個人情報保護委員会「個人情報の保護に関する法律についてのガイドライン（通則編）」3-6-3、及び「『個人情報の保護に関する法律についてのガイドライン』に関するQ&A」7-43。

人が同意した目的の範囲内であっても生じうる差別的取扱い、不当な格差、行動操作等に対しては、同意取得という手続的規律だけでは実質的に対応できない。現行法は、令和2年の改正により、不適正利用の禁止義務²⁷を導入したが、「違法又は不当な行為を助長し、又は誘発するおそれがある方法により個人情報を利用してはならない」とその要件は抽象的である。ガイドラインでは「法（個人情報保護に関する法律）その他の法令に違反する行為、及び直ちに違法とはいえないものの、法（個人情報保護に関する法律）その他の法令の制度趣旨又は公序良俗に反する等、社会通念上適正とは認められない行為」²⁸を禁止するものであるとし、その典型例を示すが、具体的な考慮要素が示されているわけではない。このように、現行法は「処理の入口」（同意に代表される本人の関与）を重視する一方、「処理の内容・結果がもたらすリスク」への規律が薄いという構造的な問題を抱えている。

これら二つの課題に関連して、個人情報の要件である「識別可能性」の境界が技術の進歩によって流動化しているという問題も法益保護の観点から無視できない問題である。従来の個人情報保護法制は、特定の個人を識別できる情報（個人情報）と識別できない情報（統計情報・匿名加工情報等）とを区別し、後者には厳格な規制を課さないというアーキテクチャを基本としてきた。しかし、機械学習・外部データとの突合・再識別技術の高度化により、「匿名化された」データから個人が再識別されるリスクは否定できず、「識別できない処理」と「識別できる処理」の境界はもはや技術的に安定したものではない。

（3）「手続的規律」から「実体的規律」への転換の必要性

以上の課題に対処するためには、本人同意等の「本人が関与したか否か」という手続的な問いから、「当該処理が個人の権利利益にいかなるリスクをもたらし、そのリスクを社会的に許容できる水準まで低減するためにいかなる措置が必要か」という実体的な問いへの転換が不可欠である。この転換の中核にあるのは、個人情報の取扱いによって生じるリスクの類型化と、そのリスクの程度に応じた規律の比例的設計（リスクベース・アプローチ）という発想である。

データ処理技術の発展に伴い、個人情報の処理がもたらす法益へのリスクは、以下の4類型に整理し得る。第一に、特定個人を評価・選別しその結果に基づいて不利益を与えるリスク（差別・格差の創出）。第二に、本人への到達性（ターゲティング）を利用した不当な働きかけのリスク（マニピュレーション・詐欺的勧誘）。第三に、本人が秘匿したい情報が第三者に知られるリスク（プライバシーの侵害）。第四に、自己の個人データを自らの意思によってコントロールできないというリスク（自律性の喪失）である。

これら4類型のリスクは、処理の技術的文脈（処理規模・自動化度・外部データとの結合可能性等）に応じてその顕在化の程度が大きく異なる。したがって、規律

²⁷ 個人情報保護法第19条。

²⁸ 個人情報保護委員会「個人情報の保護に関する法律についてのガイドライン（通則編）」3-2。

の強度もリスクの程度に応じて設計されるべきであり、「全ての処理に同一の手続的規律を適用する」という現行法の設計は、「過剰規制」と「規制漏れ」を同時に生み出す根本的な原因となっている。

その上で、リスクの程度に応じた規律の比例的設計をするためには、各リスク類型について「どの程度のリスクを社会的に受容するか」という判断が不可避となる。事例1（インフラ点検）における欠陥検出率や、事例2（自動運転）における「慎重かつ有能な人間の運転者」のように、比較的わかりやすい安全水準の参照点が存在する領域とは異なり、個人情報保護においては、たとえばAIによるプロファイリングがもたらす差別リスクをどの程度まで許容するかという問いは、純粋に技術的な問いではなく、社会がいかなる公正さの概念を採用するかという規範的な問いである。したがって、以下で論じる技術の特定（3.3.3）及びルール設計（3.3.4）は、このような法益リスクの受容可能性についての社会的合意形成を前提として初めて実効的に機能するものであることに留意が必要である。

3.3.3 ルールの機能を代替・補完できる技術（PETs）の特定

個人情報の利活用により生じ得るリスクを、社会的に受容可能なレベルにまで低減させるには、ルールとしての「制度」だけでなく、「技術」的な措置を講じることが重要である。個人情報の利活用に伴う法益侵害リスクを社会的に受容可能なレベルに低減できる「技術」であれば、「制度」が法益保護において果たしている機能を担保し、またこれを侵害しようとする行為を防止する役割を担うから、「技術」は「制度」による法的保護の実効性を高める関係にある²⁹。このような、制度的な個人情報保護規律を補完・強化する技術的手段として注目されるのが、いわゆるプライバシー保護技術（PETs: Privacy-Enhancing Technologies）である。

（1）PETsの国際的な位置づけ

OECD（経済協力開発機構）は、プライバシー保護技術（PETs: Privacy-Enhancing Technologies）を、制度的な個人情報保護規律を補完・強化する技術的手段として体系的に整理しており、2023年のOECDレポート「Emerging Privacy-Enhancing Technologies」³⁰において、PETsを（i）データの難読化（匿名化・仮名化・差分プライバシー等）、（ii）暗号化データ処理（秘密計算・準同型暗号・マルチパーティ計算等）、（iii）連合型・分散型分析（連合学習・分散型分析等）、（iv）説明責任ツール（監査可能ログ・透明性向上技術等）の4つに整理している。

OECDレポートでは、PETsそのものを法的な枠組みに代わるものとして位置づけるのではなく、法制度が要求する保護水準を技術的に達成するための手段として

²⁹ デジタル庁「データセキュリティワーキンググループとりまとめ」（2026年3月11日）6頁。

³⁰ OECD (2023), "Emerging privacy-enhancing technologies: Current regulatory and policy approaches", OECD Digital Economy Papers, No. 351, OECD Publishing, Paris. <https://doi.org/10.1787/bf121be4-en>

位置づけている。すなわち、PETsは制度的義務の「代替」ではなく「実装手段」であり、PETsを用いることによってリスクが十分に低減された場合には、より緩やかな制度的規律で足りるという比例原則的な発想と結びついていると評価できる。

同時に、OECDレポートではPETsが万能薬ではないことを明示している。PETsはデータ処理によるバイアスの問題を解決するものではなく、また処理を行うITシステム全体の安全性を保障するものでもない。データガバナンスを実現するためには、PETsの活用に加えて、事業者内部の規律・透明性の確保・プライバシー影響評価の実施・規制当局による監督という組織的・制度的ガバナンスが不可欠であることが強調されている。この「技術と制度・運用の組み合わせ」という観点は、LE4SDSの統合的設計の枠組みと一致する。

（2）リスク類型ごとのPETs

以下では、OECDの4類型に対応しつつ、3.3.2（3）で整理した4つのリスク類型との関係でPETsを整理する。なお、以下に例示する各技術は、それぞれ単独で完結するものではなく、複数の技術と制度的ガバナンスとを組み合わせる用いることが前提となる。

ここで重要なのは、PETsの役割が、リスクの低い処理にのみ用いられる付加的な措置ではなく、むしろリスクの高い処理を社会的に受容可能な水準に引き下げるための中核的手段として位置づけられるという点である。匿名化や仮名化によってそもそも識別性が排除された処理は、それ自体がリスクの低い処理であり、緩やかな制度的規律で足りる。これに対して、個人を識別可能な状態での処理が不可避である場面（たとえば、医療データの横断的分析や金融取引の不正検知等）においてこそ、秘密計算や連合学習といった高度なPETsが、処理の有用性を維持しつつリスクを技術的に低減する手段として機能する。

（a）評価・選別による不利益（差別格差の創出）リスクへの対応

AIによる処理を含む個人情報を用いた評価・選別から生じるリスクに対しては、第一に、説明可能AI（XAI: Explainable AI）が有効である。AIによる評価・判断のプロセスと根拠を人間が理解できる形で可視化する技術であり、不当・不公平な評価が行われていないかを検証・是正する機会を生む。第二に、公平性AI（Fairness AI）として、AIモデルが特定の属性（性別・人種・出身地等）によって偏った判断をしないよう設計・評価する手法がある。第三に、差分プライバシーは、データセットに統計的なノイズを加えることで特定個人のデータが分析結果に与える影響を極小化し、個人を狙い撃ちした評価を技術的に防止する。ただし、これらの技術は差別・偏見のリスクを軽減する補助的手段であり、アルゴリズムの社会的公正性に関する規範的な判断を代替するものではない。

(b) 本人到達性（ターゲティング）の濫用リスク（マニピュレーション・詐欺的勧誘）への対応

ターゲティングによる不当な働きかけのリスクに対しては、個人の特定を困難にするデータ加工技術が有効である。**k**-匿名化（同一属性を持つ個人が**k**人以上存在する状態にする技術）、**l**-多様性（匿名集合内の機微情報の多様性を確保する技術）、**t**-近接性（機微属性の分布を保存する技術）等の統計的匿名化手法は、個人の特定・到達を技術的に困難にすることで本人へのターゲティングリスクを軽減する。また、仮名化（氏名等の直接識別子を仮IDに置換する処理）は、データの有用性を維持しつつ直接識別のリスクを低減する。

(c) 私生活暴露リスク（プライバシーの侵害）への対応

本人の私的領域が第三者に知られるリスクに対しては、暗号化処理とデータ分散化が主要な技術的手段となる。秘密計算（準同型暗号・マルチパーティ計算）は、データを暗号化したまま統計分析・機械学習等の計算を行う技術であり、処理者自身がデータの内容を知ることなく計算を行うことを可能にする。連合学習（**Federated Learning**）は、各デバイス・組織が保有するデータを外部に送みせず、学習モデルのみを共有・統合する技術である。エンドツーエンド暗号化は、通信・保存データを暗号化し管理者を含む第三者による覗き見を防止する。

(d) 自己情報コントロール喪失（自律性の喪失）リスクへの対応

本人が自らのデータの利用状況を把握・制御できないリスクに対しては、制度的手段と技術的手段の組み合わせが有効である。パーソナルデータストア（**PDS**）・情報銀行は、個人データを本人管理下の領域に集約し、本人の明確な意思に基づいて第三者提供を許可する仕組みであり、データ利用の主導権を個人が握ることを可能にする。同意管理プラットフォーム（**CMP**）は、利用目的・提供先ごとに細かく同意・撤回を管理できるツールである。また、透明性向上技術（**TETs**）は、データの利用履歴を可視化するダッシュボードを通じて本人が実際の利用状況を確認できる環境を整える。

以上を総括すると、**LE4SDS**の観点からは、現行法の手続的規律が果たしている「エージェンシー問題の解決」という機能は、**PETs**によって以下のように技術的に代替・補完し得る。すなわち、同意取得による本人意思の確認は、アクセス管理・秘密計算等によるデータの目的外利用の技術的防止によって補完される。匿名加工による識別性の排除は、差分プライバシー・**k**-匿名化等によるより精緻な再識別リスクの低減によって補完される。そして、利用目的の通知による透明性の確保は、監査可能ログ・**TETs**によるデータ利用状況の技術的可視化によって補完される。ただし、これらの技術はいずれも万能ではなく、技術的措置と制度的ガバナンスの組み合わせによって初めて実効性が担保されることは、次節で論じるとおりである。

3.3.4 技術の活用を促すためのルール及び運用方法の設計

前節で特定した PETs の活用を制度的に促進し、技術とルールの統合的運用を実現するためには、ルール及び運用方法の設計が必要となる。本節では、まずこの点に関する国際的な制度設計の先行事例を分析し（（1）（2））、その上で我が国における統合的設計の論点を整理する（（3））。

（1）GDPR の設計思想：実体的規律の先行モデル

現行の個人情報保護制度をめぐる国際的な議論において、最も影響力を持つ規範体系は EU の一般データ保護規則（GDPR³¹）である。GDPR は、個人の同意を処理の正当化根拠の一つとして位置づけながらも、同意（a）を含む 6 つの適法性根拠³²を列挙し、同意以外の根拠によっても処理が正当化される設計ではない。

GDPR が実体的規律を採用していることは随所に見られる。たとえば、GDPR が定める個人情報の取扱いと関連する基本原則³³としての、目的の限定やデータ最小化の原則が挙げられる。また、個人情報の処理の正当化根拠の 1 つである「正当な利益（legitimate interests）」は、比例原則を要求するものである。さらに、高リスクな処理を行う前に、想定されるリスクの性質・範囲・目的・深刻度を事前に評価し、リスク低減措置の妥当性を確認することを事業者に義務付ける仕組みである「データ保護影響評価（DPIA）」³⁴により、「何をするか」ではなく「いかなるリスクを生み出し、いかに低減するか」が規律の中心に置かれている。

また、GDPR は「適切な技術的・組織的措置（TOMs: Technical and Organisational Measures）」の実装を管理者に義務付けており³⁵、その具体的内容はリスクの性質・範囲・深刻度に応じて決定されることとされている。すなわち、明文上、「技術的措置」を実装することを求め、技術的措置の選択と実装をリスクの程度に連動させ、「どのような技術を使うべきか」は事業者がリスク評価に基づいて決定するという設計である。この点は、特定の技術を規格として一律に指定するシンガポールのファサード画像診断 AI のアプローチとは対照的な、原則ベースの規制設計といえる。

GDPR はさらに、「データ保護バイデザイン（Data Protection by Design）」および「データ保護バイデフォルト（Data protection by Default）」の原則を明文化している³⁶。これは、システムの設計段階からデータ保護を組み込むことを事業者

³¹ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation)

³² GDPR, Art. 6(1). 適法性の根拠として (a) 本人の同意、(b) 契約の履行、(c) 法的義務の遵守、(d) 生命に関わる利益の保護、(e) 公共の利益又は公的権限の行使、(f) 管理者又は第三者の正当な利益、の 6 つを列挙する。

³³ GDPR, Art. 5

³⁴ GDPR, Art. 35.

³⁵ GDPR, Arts. 24, 25, 32.

³⁶ GDPR, Art. 25. 同条第 1 項が Data Protection by Design を、第 2 項が Data Protection by Default を規定する。

義務付けるものであり、技術と制度の統合的設計という LE4SDS の発想と方向を同じくするアプローチである。

（２）日本の改正動向とその限界

我が国の個人情報保護法も、いわゆる「3年ごと見直し」³⁷において、現行法の過剰規制に対処する方向での改正が検討されている。具体的には、複数の事業者が保有するデータを横断的に統合して統計情報等を作成するニーズへの対応として、統計情報等の作成にのみ利用されることが担保されている場合に限り、本人同意なしで第三者提供することや要配慮個人情報を取得することを認める方向が示されている。

この改正の方向性は、「本人の権利利益への直接の影響の有無」を軸に同意規制の適用範囲を整理し直すものであり、比例原則的なアプローチへの漸進的な転換として評価できる。ただし、同意を不要とするための担保条件として検討されているのは、（i）一定の事項の公表（提供元・提供先の氏名・名称、統計情報等の作成の内容）、（ii）「統計情報等の作成のみを目的とした提供」である旨の書面による提供元・提供先間の合意、（iii）取得者及び提供先による目的外利用・第三者提供の禁止、という人間によるモニタリングと契約的規律に依拠した措置にとどまっている。

この点に、現行法を基礎とする改正議論の限界があるように思われる。個人情報保護委員会による違反事実の検知の困難さや、昨今、委託先による違法な個人データの取扱い事案が散見されることを踏まえると、前記の担保条件が本人の権利利益の実質的な保護として十分といえるか、ひいては、データ活用のための前提としての信頼の確保足り得るか疑問が残る。

「人間が完全にコントロールできる」ことを前提にした制度は限界があり、法益保護の実効性を補完する技術的措置の要求が不可欠である。この点こそが LE4SDS の「③技術の活用を促すためのルール及び運用方法の設計」の中核をなす。

（３）「ルール×技術×運用」統合運用上の論点

以上のとおり、個人情報の処理に伴う法益侵害のリスクを低減するための規律である「ルール」は、法益保護の実効性を高めるため、「技術」による担保が不可欠である。また、「ルール」及び「技術」を実効的に機能させるためには、個人情報の処理に問題が生じた場合には迅速に対処できるガバナンス体制を構築し、現場を統制する「運用」が不可欠である。このように、個人情報の利活用に伴う法益侵害リスクを、社会的に受容可能なレベルに低減するためには、「ルール」「技術」

³⁷ 個人情報保護委員会「個人情報保護法 いわゆる 3年ごと見直しの制度改正方針」（令和 8 年 1 月）。https://www.ppc.go.jp/files/pdf/01-1_seidokaiseihousin.pdf

「運用」の三位一体の措置を設計し講じることが重要である³⁸。以下では「ルール×技術×運用」を統合的に設計するうえでの主要な論点を四点整理する。

第一に、法益リスクの受容可能性に関する社会的合意形成プロセスの制度化である。3.3.2 (3) で指摘したとおり、個人情報保護における法益は多義的であり、それぞれのリスクの受容可能性は技術的文脈のみならず社会的・文化的な価値観に依存する。したがって、リスク連動型の規律設計が実効的に機能するためには、規制当局・事業者・技術者・市民社会・学術機関が参画するマルチステークホルダー型の継続的な対話の場を制度的に設け、リスクの受容可能性とそれに対応する技術的・制度的措置の組み合わせについて、動的に合意形成を図っていく仕組みが必要である。この点は、事例1や事例2においても性能基準や安全基準の策定・更新にあたって官民協働が必要とされたのと共通するが、保護法益自体が多義的で価値判断を内包する個人情報保護の領域においては、合意形成プロセスの制度化がとりわけ本質的な意味を持つ。

第二に、技術的措置が制度的義務を代替できる条件の明確化である。PETsを実装することによって制度的規律が緩和される「条件」は、現行法においても改正案においても明確化されていない。GDPRにおけるTOMsのように、リスク評価に基づいて技術的措置の内容を事業者が決定し、その妥当性を規制当局が審査するという仕組みを設けることが重要であり、技術的措置の「種類」ではなく「達成される保護水準」を基準とすることが原則ベース規制の本旨に合致する。

第三に、組織的ガバナンスとの組み合わせの必要性である。OECDレポートが指摘するとおり、PETsはデータガバナンス全体の一部を構成するにすぎない。PETsの実装と並行して、事業者内部における規律（プライバシー・バイ・デザインの実践）、PIAの定期的実施、規制当局への透明性の確保、および違反発生時の責任の明確化という組織的ガバナンスが機能して初めて、技術的措置の実効性が担保される。

第四に、規制当局の技術評価能力の整備である。事業者が選択したPETsの適切性・有効性を規制当局が審査するためには、当局自身が技術を評価する専門的能力を持つ必要がある。事例1（シンガポール）においてAI評価機関がAIシステムの事前審査を担ったように、個人情報保護委員会においても技術的評価能力の組織的整備が急務となっている。

3.3.5 我が国の制度設計への示唆

以上の分析から、我が国の個人情報保護制度の設計に向けた示唆として以下の点が導かれる。

第一に、手続的規律から実体的規律への段階的転換である。3.3.2で確認したとおり、現行法の手続的規律（同意ベース規律）は、過剰規制と規制漏れという二つの構造的問題を抱えている。GDPRが示すように、処理のリスクの性質・程度に応じ

³⁸ デジタル庁「データセキュリティワーキンググループとりまとめ」（2026年3月11日）6頁。

て規律の強度を比例的に設計するリスクベース・アプローチへの転換が求められる。その際、リスクの低い処理（匿名加工・仮名加工等により識別性が排除又は大幅に低減された処理）については緩やかな制度的規律で足りるとし、個人を識別可能な状態での処理については高度な **PETs** の組み合わせ実装を制度的に促し、さらに処理の結果が本人の法的地位又は重要な生活上の利益に直接かつ重大な影響を及ぼし得る処理（与信判断、保険引受の可否、採用選考等における個人の評価・選別を伴う処理や、要配慮個人情報の大規模処理等）については事前の **DPIA**・処理の制限・継続モニタリングを求めるといふ、段階的な規律設計が検討に値する。

第二に、**PETs** の活用を促す制度的インセンティブの整備である。3.3.3 で特定した **PETs** が現行法のルールの機能を代替・補完し得ることは示されたが、現行法の枠組みでは **PETs** を実装することによる制度的な「見返り」（制度的規律の緩和等）が明確ではない。技術的措置の実装と制度的規律の緩和との間の条件を明確化し、事業者が **PETs** を積極的に導入するインセンティブを制度的に設計することが重要である。

第三に、法益リスクの受容可能性に関する社会的対話の場の制度化である。3.3.2 (3) で指摘したように、個人情報保護における「何がリスクであり何が許容されるか」という問いは、技術的に一義的に決定できない規範的問いである。この点は、事例1・2とは決定的に異なる、個人情報保護に固有の課題である。規制当局・事業者・技術者・市民社会が参画するマルチステークホルダー型の対話を制度的に組み込み、リスクの受容可能性についての継続的な合意形成と更新を図る仕組みが不可欠である。

第四に、規制当局の技術評価能力の確保である。原則ベースの規制設計が実効的に機能するためには、事業者が選択・実装した技術的措置の保護水準を、規制当局が専門的に審査できなければならない。個人情報保護委員会における技術専門人材の確保・育成、及び外部の技術専門機関との連携体制の構築が急務である。

4. 提言：LE4SDS の実装に向けた制度設計の方向性

4.1 総論：三事例から得られる共通の知見

本研究は、**Software-Defined Society (SDS)** における法システムの在り方として、**LE4SDS (Legal Engineering for Software-Defined Society)** という枠組みを提示してきた。その核心は、「法益保護」という法の本来の目的を実現するにあたって、人間に向けた行為規範としての「ルール」と、リスクの評価・低減・管理を可能にする「技術」とを、対等な手段として統合的に運用するという発想にある。

第3章では、この枠組みを三つの事例に適用して分析を行った。インフラ点検（事例1）、自動運転車（事例2）、及び個人情報保護（事例3）である。これら三事例を横断して浮かび上がる共通の知見は、従来の「事前に人や組織が遵守すべき具体的な行為規範を定め、その遵守をもって適法性を確保する」という規制プロ

一ちが、技術の急速な進歩と社会の複雑化に伴い、構造的な限界に直面しているということである。

しかし、この共通の知見の背後には、事例ごとに質的に異なる課題が存在する。事例1では、検出すべき欠陥の種別が列挙可能であり、入力データの形式や可否基準を数値で規格化できるという条件のもとで、定量的な性能基準の設定と評価プロセスの制度化が主たる課題であった。事例2では、入力空間とリスクシナリオが事実上無限であるために性能基準の事前網羅が困難であるという、より本質的な制約のもとで、動的な安全保証と責任帰属の再設計が求められた。そして事例3では、点検や運転といった特定のタスクの代替ではなく、プライバシーという法益そのものが規律の対象であり、保護法益自体が多義的でリスクが処理の技術的文脈に依存するという、さらに根源的な問題が提示された。

これらの事例ごとの課題の違いは、1.2.2で整理した従来型規制の三つの限界——①コンプライアンスの具体的方法を事前に定義することの限界、②ある時点での規範遵守によって適法性を確保することの限界、③達成すべき結果（Outcome）自体を定義することの限界——と対応している。事例1（類型A）では、タスクの限定性ゆえに①コンプライアンスの方法も②その評価基準も事前に定義可能であり、従来型の限界を比較的克服しやすい。事例2（類型B）では、入力空間の無制限性ゆえに①②がともに困難であることに加え、③「慎重かつ有能な人間の運転者」という安全水準の参照点を定量的に定義すること自体が困難であるという、より根本的な課題に直面する。そして事例3（類型C）では、そもそも③何をもって「許容可能なリスク」とするかという結果の定義が規範的・社会的な問いとならざるを得ず、それゆえに①②の設計もまた困難となるという、三つの限界が重層的に現れる領域である。

以下では、この三事例の分析から得られた知見を、LE4SDSの実装に向けた具体的な提言として整理する。まず、リスク特性に応じた規律設計の類型化を行い

（4.2）、次に各類型に固有の制度設計課題を論じ（4.3）、最後に三類型に共通して必要となる制度基盤の整備について提言する（4.4）。

4.2 リスク特性に応じた規律設計の三類型

三事例の分析から、技術が代替・補完する対象のリスク特性に応じて、LE4SDSの規律設計は以下の三つの類型に整理できる。この類型化は、ある領域においてルールと技術の統合的運用を制度化しようとする際に、「当該領域のリスク特性がいかなるものであるか」を出発点として、適切な規律設計のアプローチを選択するための指針を提供するものである。

類型	リスク特性	規律設計のアプローチ	参照事例
類型 A	タスクが限定的で、リスクシナリオが列挙可能。入力データの形式・品質要件が規格化でき、合否基準を数値で表現できる。	定量的性能基準の事前規格化+承認制	インフラ点検（シンガポール TR 78 シリーズ）
類型 B	タスクが複雑かつ入力空間が無限で、リスクシナリオの事前網羅が困難。性能基準の網羅的な事前規格化に原理的限界がある。	安全原則+安全ケースによる事前評価と、ライフサイクル型の継続的モニタリングの組み合わせ	自動運転車（英国 Automated Vehicles Act 2024）
類型 C	特定のタスクの代替ではなく法益保護そのものが規律対象。保護法益自体が多義的で、リスクが処理の技術的文脈に依存する。	法益リスクの受容可能性に関する社会的合意形成を前提とした、リスク連動型の段階的規律設計	個人情報保護（GDPR、OECD PETS レポート、日本の改正動向）

この三類型は排他的なカテゴリーではなく、連続的なスペクトラムとして理解されるべきである。ある領域が技術の発展に伴って類型 A から類型 B へと移行する場合もあり得るし（たとえば、インフラ点検において AI が有資格者を上回る性能を実証し、**Human-in-the-loop** を前提としない制度設計が求められる段階に至った場合）、一つの領域が複数の類型の特性を併せ持つ場合もある。重要なのは、規律設計の出発点として、対象領域のリスク特性を正確に把握することである。

4.3 各類型に固有の制度設計課題

4.3.1 類型 A における課題：性能基準の設定と評価プロセスの制度化

類型 A の強みは、タスクの限定性ゆえに定量的な性能基準の設定が可能であり、客観性と再現性の高い評価プロセスを構築できる点にある。シンガポールの TR 78 シリーズが示したように、欠陥種別ごとの検出再現率、匿名化処理の達成率、撮影画像の最低分解能といった定量的基準を規格として明確化し、AI 評価機関による事前評価と定期的再評価を組み合わせることで、技術代替の社会的信頼性を確保することができる。

我が国のアナログ規制撤廃は、LE4SDS の三段階のうち「①ルールが果たしている機能の特定」及び「②ルールの機能を代替・補完できる技術の特定」を部分的に遂行したものと評価できる。しかし、「③技術の活用を促すためのルール及び運用方法の設計」については、ほとんどの分野で手つかずの状態にある。すなわち、コ

ンプライアンス目的で使用される技術の性能要件が明確化されておらず、具体的にどのような性能のシステムを導入すべきかは事業者の判断に委ねられている。

このアプローチは、自ら適切なリスク判断ができる事業者にとっては柔軟で使いやすい制度である。しかし、技術の性能評価に関する専門的知見が限られた事業者にとっては、新技術導入に際しての法的リスクの不透明さが障壁となり、結果として技術活用が進まないおそれがある。

したがって、類型Aに該当する領域については、以下の制度整備を提言する。第一に、法益保護に資する技術について、分野ごとに定量的な性能基準を規格として策定することである。その際、基準の策定にあたっては、規制当局と技術専門家及び事業者の協働が不可欠である。第二に、性能基準への適合性を客観的に評価するための評価プロセス（評価手法・テストデータの形式・評価機関の要件等）を制度化することである。第三に、技術の進歩に対応するため、性能基準と評価プロセスの定期的な見直しの仕組みを組み込むことである。

4.3.2 類型Bにおける課題：動的な安全保証と責任帰属の再設計

類型Bにおいては、入力空間とリスクシナリオが事実上無限であるため、類型Aのように性能基準を事前に網羅的に規格化することには原理的な限界がある。英国のAVAが示したように、この類型では、事前の型式承認と事後の継続的モニタリングを組み合わせたライフサイクル型の安全保証が不可欠となる。

第一の課題は、事前評価と事後モニタリングの制度的統合である。AVAは、自動運転車に対する「安全原則声明（SoSP）」に基づく事前の自己走行テストと、稼働後のin-use assurance制度を法律上の二本柱として位置づけた。事前評価においては、安全ケースアプローチにより、事業者が自らの技術と運用領域（ODD）の特性に応じた評価根拠を規制当局に提示する責任を負う。事後モニタリングにおいては、先行指標（リスク事象の検出・ハザード報告）と遅行指標（衝突・死傷者数等）の双方を用いた継続的な安全パフォーマンス評価が義務付けられる。我が国においても、類型Bに該当する領域では、事前の認証のみに依拠するのではなく、稼働後のデータ収集・モニタリング・定期的再評価を法制度上の義務として組み込むことを提言する。

第二の課題は、責任帰属の枠組みの再設計である。自動運転のように、人間の個別の意思に基づかない形でシステムが挙動を決定する領域では、従来の「誰の何の過失か」という問いに原理的に答えられない場面が生じる。我が国においては過失及び因果関係が比較的広く認められる傾向にあり、開発者個人への過失犯としての訴追リスクが技術開発の萎縮要因となりかねない。AVAは、ASDE（Authorised Self-Driving Entity）という組織的な安全確保責任の帰属主体を法的に創設することで、個人の過失責任への依拠から組織としての継続的な安全確保義務の履行へと、責任構成を転換した。我が国においても、類型Bに該当する領域では、開発者・製造者・運行事業者のいずれがいかなる安全確保義務を継続的に担うかを制度的に明

確化し、個人の過失責任への過度な依拠を避ける方向での制度設計を検討すべきである。

第三の課題は、誠実な情報開示を促すインセンティブ構造の設計である。ライフサイクル型安全保証が実効的に機能するためには、事業者が走行ログ・インシデントデータ等を誠実に開示することが大前提となる。しかし、インシデント情報の積極的な開示は法的・競争的リスクを高める可能性があり、隠蔽・過小報告への誘因が生じやすい。AVAは、法人と個人（指名管理者・関連上級管理職）の双方に対する刑事制裁と、デュー・ディリジェンス・ディフェンス（義務違反を避けるための合理的な予防措置を適切に実施したことの立証による免責）を組み合わせることで、誠実に情報を開示する事業者が法的保護を受け、隠蔽を図る事業者が制裁を受けるという非対称的な誘因構造を制度化した。このような情報開示インセンティブの設計は、類型Bにおける安全文化の構築に不可欠であり、我が国においても検討に値する。

4.3.3 類型Cにおける課題：法益リスクの受容可能性に関する社会的合意の形成

類型Cが類型A及びBと決定的に異なるのは、技術と制度の統合的設計に先立って、「何をもって許容可能なリスクとするか」という問い自体に対する社会的合意が必要であるという点にある。

類型Aにおいては、欠陥検出率や匿名化処理達成率のように、性能基準の設定に際して比較的明確な参照点が存在する。類型Bにおいても、「慎重かつ有能な人間の運転者（C&C human driver）」という、定量化は困難であるものの概念的には共有可能な安全水準の参照点がある。これに対して類型Cでは、保護法益自体が多義的であり（差別格差の創出、マニピュレーション、プライバシーの侵害、自律性の喪失等）、それぞれのリスクの受容可能性が、処理の技術的文脈（処理規模、自動化度、外部データとの結合可能性等）のみならず、社会的・文化的な価値観にも依存して変動する。たとえば、AIによるプロファイリングがもたらす差別リスクをどの程度まで許容するかは、純粋に技術的な問いではなく、社会がいかなる公正さの概念を採用するかという規範的な問いである。

したがって、類型Cにおいては、リスク連動型の段階的規律設計（技術的担保措置の強度と制度的規律の強度を、リスクの程度に応じて比例的に設定する設計）を行うにあたって、その前提として、法益リスクの受容可能性についての継続的な社会的合意形成プロセスを制度的に組み込むことが不可欠である。

具体的には、以下の三段階のプロセスが求められる。

第一に、法益リスクの類型化と可視化である。3.3.2 (3) で整理したように、個人情報取扱いにより生じるリスクを、評価・選別による不利益、ターゲティングの濫用、私生活の暴露、自己情報コントロールの喪失といった類型に整理し、それぞれのリスクがいかなる技術的文脈において顕在化するかを、具体的な事例とともに社会に対して可視化することが出発点となる。

第二に、リスクの受容可能性に関するマルチステークホルダー型の対話プロセスの制度化である。規制当局・事業者・技術者・市民社会・学術機関が参画する継続的な対話の場を設け、各リスク類型について、社会的に受容可能な水準と、その水準を達成するために求められる技術的・制度的措置の組み合わせについて、合意形成を図ることが必要である。この合意は一度限りのものではなく、技術の発展や社会的価値観の変化に応じて動的に更新されなければならない。

第三に、合意に基づくリスク連動型の規律設計の実装である。3.3.5で示したように、リスク水準に応じて、たとえば、低リスク処理（匿名加工・仮名加工等により識別性が排除又は大幅に低減された処理）には緩やかな事後監督を、中リスク処理（個人を識別可能な状態での処理であるが処理目的・範囲が限定的なもの）にはPIAの実施と規制当局への報告を、高リスク処理（処理の結果が本人の法的地位又は重要な生活上の利益に直接かつ重大な影響を及ぼし得るもの）には事前のDPIA・処理の制限・継続モニタリングを、それぞれ求める段階的な規律を制度化する。その際、技術的措置が制度的義務を代替できる条件（いかなるPETsをいかなる方法で実装すれば、いかなる制度的規律が緩和されるか）を、上記の合意形成プロセスの中で明確化していくことが重要である。

類型Cにおけるこのような社会的合意形成プロセスの制度化は、単に個人情報保護の分野のみならず、AIの社会実装がもたらす倫理的・社会的リスクへの対処全般に通じる課題でもある。「リスクの受容可能性」という問いが技術的に一義的に決定できない領域では、マルチステークホルダーによる継続的な対話と合意の更新が、LE4SDSにおける「③技術の活用を促すためのルール及び運用方法の設計」の中核をなすのである。

4.4 三類型に共通する制度基盤の整備

以上の各類型に固有の課題に加えて、LE4SDSの実装を支える横断的な制度基盤として、以下の四点の整備が重要である。

4.4.1 適合性評価の動態化

三事例に共通して浮かび上がったのは、事前の静的な適合性審査のみでは、SDSにおける技術システムの安全性・適切性を担保できないという認識である。技術や運用環境が継続的に変化する以上、ある時点での適合性評価は急速に陳腐化する。

この課題に対して、三類型はそれぞれ異なる形で「適合性評価の動態化」を求めている。類型Aでは、定量的な性能基準の定期的更新と再評価（シンガポールTR 78-2が推奨する2年以内ごとの再評価）が中心となる。類型Bでは、事前の型式認証に加えて稼働後の継続的モニタリング（AVAのin-use assurance）が法制度上の柱として位置づけられる。類型Cでは、プライバシー影響評価（PIA/DPIA）がデータ処理のライフサイクルにわたって継続的に実施されることが求められる。

我が国においても、規制の対象となる技術システムについて、事前評価・事後モニタリング・定期的再評価を一体的に運用する動的な適合性評価の枠組みを、各分野の特性に応じて制度化していくことが不可欠である。

4.4.2 規制当局の技術評価能力の整備

LE4SDSが原則ベースの規制設計を志向する以上、規制当局は、事業者が選択した技術的措置の適切性・有効性を実質的に審査する能力を備えなければならない。類型Aでは、AI評価機関がAIシステムの事前審査を担う体制が必要である。類型Bでは、安全ケースの妥当性を評価し、稼働後のモニタリングデータを分析する能力が求められる。類型Cでは、事業者が実装したPETsの保護水準を検証し、リスク評価の妥当性を審査する専門的知見が不可欠である。

現状では、我が国の規制当局がこれらの技術評価能力を十分に備えているとは言い難い。この課題に対処するためには、少なくとも以下の三つの方策を組み合わせる必要がある。第一に、規制当局内部における技術専門人材の確保・育成である。これには、技術系バックグラウンドを持つ人材の採用のみならず、既存の法律専門職員に対する技術リテラシー教育の体系的な実施も含まれる。第二に、外部の技術専門機関との連携体制の構築である。シンガポールのTR 78-2がAI評価機関（AI assessor）の制度を設けたように、技術評価の一部を規制当局の指定する外部専門機関に委託する枠組みが有効であり得る。第三に、規制当局自身がデジタル技術を活用して規制業務を高度化すること、たとえばモニタリングデータの自動分析やリスク事象の自動検出といった技術の導入により、限られた人的リソースの中でも実効的な審査を可能にする体制を構築することである。これらの取組なくしては、原則ベースの規制設計は「形式的な原則の提示」に終わり、その実効性は確保されない。

4.4.3 マルチステークホルダー型ガバナンスの制度化

LE4SDSが「ルールと技術の統合的運用が最先端の技術動向を踏まえてタイムリーに更新される」ことを本質的要素とする以上、制度を動的にアップデートし続けるための仕組みが不可欠である。

マルチステークホルダー型のガバナンスは、三類型のいずれにおいても必要であるが、その機能は類型ごとに異なる。類型Aでは、性能基準と評価プロセスの技術的妥当性を継続的に検証し更新するための専門的対話の場として機能する。類型Bでは、安全基準が世界的にも未確定である領域において、技術・法制度・運用の三者が継続的に対話しながら制度を動的にアップデートしていくための基盤となる。類型Cでは、法益リスクの受容可能性に関する社会的合意形成のプロセスそのものを担う。

我が国においても、政府・事業者・技術者・学術機関・市民社会が同等の立場で継続的に意見交換を行う常設の対話プラットフォームを、分野横断的に設立するこ

とが重要である。このようなプラットフォームは、LE4SDSにおける「③技術の活用を促すためのルール及び運用方法の設計」をアジャイルに更新し続けるための不可欠の制度的基盤となる。

なお、マルチステークホルダー型の対話プラットフォームに対しては、参加者間の情報格差・交渉力格差が実質的な対話を困難にするのではないか、あるいは合意形成の遅延が技術の急速な発展に追いつかないのではないか、といった批判が想定される。この点について、アジャイルガバナンスの本質は、完璧な合意を待つことではなく、暫定的な合意に基づいて制度を運用し、結果のフィードバックを受けて修正するというイテレーティブなプロセスにある点を強調しておきたい。したがって、対話プラットフォームの設計においては、合意形成の速度と制度更新の頻度を意識的に組み込むことが重要であり、これを怠れば常設の場が形骸化するリスクがある。

4.4.4 国際的な基準調和への戦略的参画

LE4SDSの実装は、一国の制度設計の中で完結するものではない。技術が国境を越えて展開される以上、安全基準・性能評価手法・データ保護水準の国際的な調和なくしては、事業者の国際競争力の確保も、市民の保護水準の維持も困難である。

類型Bに関しては、英国がUNECEにおいて自動運転システムに関する国際型式認証規則の策定を主導しているように、各国の規律設計が国際的な基準調和の枠組みの中で発展させられつつある。類型Cに関しても、OECDにおけるPETsの体系的整理やGDPRのTOMsの概念が、技術と制度の統合に向けた国際的な議論の基盤を提供している。

我が国は、特定の規制モデルを一方向的に輸出・輸入するのではなく、各国・各地域の制度がそれぞれの社会的文脈に応じて多様であることを前提としつつ、相互運用性を確保するという立場から、国際的な基準調和の議論に積極的に参画すべきである。LE4SDSの「法益保護を実現するためにルールと技術を統合的に運用する」という枠組みは、特定の規制哲学に依拠しない機能的なアプローチであるがゆえに、多様な法制度間の共通言語となり得る可能性を持っている。

4.5 結語

本研究が提示するLE4SDSの核心は、「法とは、ルールと技術の統合的運用である」というテーゼにある。従来、法益保護の手段は主として人間に向けた行為規範として設計されてきた。しかし、Software-Defined Societyにおいては、技術そのものが法益保護の実現に直接貢献する手段となる。このとき、ルールの役割は、技術の活用を促し、その適切性を評価し、リスクに応じた運用方法を設計するという、より高次のガバナンス機能へと変容する。

三つの事例分析が示したのは、この変容が一様ではなく、リスク特性に応じた多様なアプローチを要するという点である。タスクが限定的な領域では定量的な性

能基準の策定が有効であり（類型A）、入力空間が無限の領域ではライフサイクル型の動的安全保証が不可欠であり（類型B）、保護法益自体が多義的な領域ではリスクの受容可能性に関する社会的合意形成が前提となる（類型C）。これらの類型化は、我が国がLE4SDSを実装するにあたって、各領域の特性に応じた規律設計の戦略を構築するための基盤を提供するものである。

SDSの到来とともに、法システムもまた、ソフトウェアと同様に、継続的に更新され、変化するリスクに俊敏に対応できるものへと進化しなければならない。その実現のためには、適合性評価の動態化、規制当局の技術評価能力の整備、マルチステークホルダー型ガバナンスの制度化、そして国際的な基準調和への戦略的参画といった、横断的な制度基盤の整備が不可欠である。LE4SDSは、このような法システムの進化の方向性を示す枠組みである。今後、我が国においてAIのもたらす正のインパクトを最大化するためには、LE4SDSの具体化に向けた学際的な研究と実践の蓄積が、ますます重要となる。

以上

「AI時代のルール（法・標準）とソフトウェアエンジニアリングに関する共同調査研究」報告書別冊1：ルールと技術の統合的運用としての「法」— **LE4SDS**（**Legal Engineering for Software-Defined Society**）の構築に向けて

京都大学大学院法学政治学研究科 法政策共同研究センター（KILAP）

*羽深宏樹（特任教授）

*北山昇（協力研究員）

稲谷龍彦（教授）

広瀬貴之（特定講師）

柴田高広（特定教授）

（* 執筆主担当）