

AI 保険の法的課題と展望

——海外動向の分析を中心に——

東京大学大学院法学政治学研究科

柳平大樹

要 旨

生成 AI の社会実装に伴い、ハルシネーション、アルゴリズムバイアス、サイバー攻撃の高度化、機密データ漏洩等のリスクが顕在化してきている。本稿では、サイバー保険の発展過程を手がかりにしつつ、AI 保険の法的課題を検討する。

サイバー保険の判例分析として、米国 Merck 判決は戦争免責条項のサイバー空間への適用という問題を提起し、G&G Oil 判決は従来型の商業犯罪保険のコンピュータ詐欺条項がランサムウェア損害に適用され得ることが示されている。これらの判例は約款の精緻化と補償範囲の明確化を促しているものと捉えられる。

海外保険市場では AI 除外条項と AI 専用保険への二極化が進む。Armilla AI/Chaucer の保険商品や Munich Re のパフォーマンス保証保険は、AI の失敗様式に適合した補償設計とデューデリジェンスを通じた引受審査を特徴としている。

日本法への含意としては、告知義務と AI インベントリの関係、AI 除外条項の解釈における約款作成者不利の原則と合理的期待の保護、内部統制システム構築義務の一内容としての AI ガバナンス体制構築義務、保険による規律付けの可能性を論じる。AI ガバナンス体制構築義務の具体的内容として、組織的ガバナンスの枠組みと AI 固有の推論段階における技術的統制の構造を考察する。

AI 保険の確立にはサイバー保険の教訓を踏まえた約款の精緻化と、AI ガバナンス体制構築義務の確立が必要であると考えられる。AI 保険は AI ガバナンス体制構築を促す規律付けの機能を担いうる。

キーワード：AI 保険、サイバー保険、ハルシネーション、サイレント AI、告知義務、保険による規律付け、AI ガバナンス、内部統制システム構築義務

1 はじめに

生成 AI の社会実装が進むにつれて従来の法的枠組みでは捕捉しきれないリスクが顕在化してきている。Stanford University の調査によると、AI インシデントの件数は 2012 年以降 26 倍に増加している¹。意思決定者の 2 人に 1 人が AI のリスクはメリットより高いと考えているとの報告もある²。こうした懸念を払拭して AI の導入を促進するには、リスク移転手段としての保険が欠かせない。しかし、日本の AI 保険市場は黎明期にあり³、保険法からの AI リスク分析も始まったばかりである。

こうした中、サイバー保険の発展過程が参考になるものとする。サイバー保険は約款の曖昧さから多くの紛争を経験し、裁判例を通じて解釈が精緻化されてきている。リスクの具体化と約款の精緻化というプロセスは、AI 保険も同様にたどることになる。サイバー保険の先行研究として、金奈穂（2019）のサイレント・サイバーリスク動向分析、牛窪賢一（2022）の米国市場動向研究、濱田和博（2022）の戦争免責条項の検討、新谷哲之介（2023）のサイバーリスクと戦争リスクの交錯に関する考察がある⁴。これらはデジタルリスクの法的性質や約款解釈に関する学説の蓄積といえる。さらに、王学士（2024）は、サイバー保険の制度化における理論パラダイムを検討し、米国で展開された「保険＝ガバナンス（Insurance as Governance）」理論に依拠して、保険者が引受決定や契約上の予

¹ Stanford University Human-Centered AI Institute (HAI), AI Index Report 2023, Chapter 3: Technical AI Ethics (2023) 9 頁 (AIAAIC データベースに基づく集計)。Munich Re, De-Risking AI Ventures: How Munich Re Assesses AI Performance Risks - Insights Into Our Due Diligence Process (Whitepaper, 2023) 8 頁, available at <https://www.munichre.com/content/dam/munichre/contentlounge/website-pieces/documents/MunichRe-De-Risking-AI-Ventures-Whitepaper.pdf> (2026 年 3 月 12 日最終閲覧) も同旨。

² Munich Re・前掲注 1) 2 頁。

³ 日本においては、現状産業界が AI の社内活用や AI ガバナンス構築という新たな課題に急ピッチで取り組んでいる状況にあり、保険の検討までは至っていないというのが現状といえるかもしれない。

⁴ 金奈穂「サイレント・サイバーリスクを巡る動向－米国・イギリスを中心に－」損保総研レポート 126 号 (2019 年) 1 頁以下、牛窪賢一「米国を中心とするサイバー保険市場の動向」損保総研レポート 138 号 (2022 年) 1 頁以下、濱田和博「国家の関与するサイバー攻撃とサイバー保険の戦争免責条項について」損保総研レポート 141 号 (2022 年) 1 頁以下、新谷哲之介「サイバーリスクと戦争リスクの交錯－ウクライナ侵攻と戦争保険の視点からの考察－」保険学雑誌 661 号 (2023 年) 125 頁、王学士「サイバー保険の理論的基礎考察(上)・(中)・(下)」NBL1264 号 (2024 年) 13 頁以下・1265 号 (2024 年) 10 頁以下・1266 号 (2024 年) 21 頁以下参照。

防措置等を通じて被保険者のリスク軽減行動を促し、第三者による監督の機能を果たしうることを論じている⁵。

また、小塚莊一郎は、AI原則の実施をコーポレートガバナンスの課題として位置づけ、サプライチェーン全体にわたる契約関係のガバナンスとして論じた⁶。同論文は、保険商品の前提条件としてAI開発者等にAI原則の実施やAIガバナンスの構築を義務づけることで、「AIリスク保険が普及するとともにAIガバナンスも関係業界に浸透するという好循環が期待できる」と述べる。本稿では、この小塚の問題提起と、王（2024）が整理した保険者の第三者による監督機能に関する視座を受け、保険法・会社法等の観点からAIガバナンスと保険の関係を具体的に検討する。

中心的な問いは、AI保険が、保険による規律付けとして機能し得るか、またそのためにいかなる基盤が必要かである。この問いに答えるため、三つの課題を設定する。まず、サイバー保険の発展過程からAI保険の約款設計にいかなる教訓が得られるか。次に、海外におけるAI保険の動向は日本にいかなる含意を持つか。そして、AI保険と内部統制システム構築義務としてのAIガバナンス体制構築義務はいかなる関係に立つか。これら三つの課題はいずれも、保険がAIガバナンスの実効的な担い手となりうるか、という一つの問いに収斂する。

本稿において、AI保険とは、AIシステムの開発・運用・利用に起因するリスクを填補の対象とする保険の総称として用いることとする。これは単一の保険商品を指すのではなく、二つの異なる形態を包含する概念である。一方は、AIリスクを積極的に補償対象として設計された専用保険である。填補の対象となる被保険利益としては、ハルシネーションや性能劣化に起因する第三者への賠償責任、AI障害による事業中断損失、規制当局対応費用等がこの類型として想定される。ここには第IV章で検討する Armilla AI/Chaucer および Munich Re の商品が該当する。他方は、サイバー保険・賠償責任保険・E&O保険等既存の保険契約においてAIリスクの補償範囲が明確でないために生じる類型である。保険者が意図せずAIリスクを引き受けている、またはその逆の状態たるサイレントAI問題がこれに該

⁵ 特に王・前掲注4）（中）12-13頁。王論文は、Baker（2010）が「第三者による監督」を国家ガバナンス理論の延長として位置づけた議論を整理し、同理論において保険者が社会リスクガバナンスの面で「準政府（quasi-government）」ないし「代理政府（surrogate government）」の性質を有すると論じられていることを紹介する。See Tom Baker, Insurance in Socio-legal Research, 6 Annu. Rev. Law Soc. Sci. 433（2010）。

⁶ 小塚莊一郎「AI原則の事業者による実施とコーポレートガバナンス」情報通信政策研究4巻2号（2021年）25頁以下。なお、本稿において「保険による規律付け」とは、保険会社がリスク評価や引受条件を通じて被保険者の行動に影響を与え、社会的に望ましい行動を促進する機能をいう（同32頁参照）。

当する。本稿は主として事業者向けの損害保険を念頭に置きつつ上記二つの問題を横断的に検討する。

II AI リスクの類型化

1 AI リスクの諸類型

AI 保険の法的課題を検討する前提として、保険契約上問題となる AI リスクを整理する。

ハルシネーションは AI 固有のリスクとして最も顕在化している⁷。生成 AI は確率論的にもっともらしい回答を生成するため、構造的に誤情報の出力を回避し難く、企業が顧客対応に生成 AI を導入した結果、存在しない商品を案内したり、法的に誤った助言を行う事態が現実には生じている。EIOPA が 2025 年に実施した EU 保険市場全域対象の生成 AI 調査でも、ハルシネーションは保険会社が認識する最大のリスクとされており、サイバーセキュリティリスク、データ保護の問題、説明可能性の欠如がこれに続く⁸。

法的責任の観点から別個に問題を提起するのがアルゴリズムバイアスである。AI モデルは学習データの偏りを反映し、特定の人種、性別や年齢層に不利な判断を下すことがある。採用、融資審査、保険引受等に AI を用いる場面では、差別訴訟リスクとして現実化している⁹。

⁷ ハルシネーションとは、生成 AI が事実に基づかない情報をあたかも真実であるかのように出力する現象をいう。

⁸ EIOPA, Generative AI Market Survey: Outlook, Use Cases and Risk Management (February 2026) 23 頁 (Figure 12)。同調査は、EU/EEA25 カ国 347 保険会社を対象とする。EIOPA 報告書によれば、回答企業は 2024 年 EU 市場の総収入保険料 (GWP) の概ね 80% に相当するが、同数値は Solvency II データベースとの照合に基づく近似値であり厳密値として扱うべきではないとされる (同報告書 6 頁脚注 6 参照)。同調査によれば、ハルシネーションリスクは、従来の教師あり機械学習には問題とならなかった生成 AI 固有の新しいリスク類型として認識されている。

⁹ 保険業界においても、アルゴリズムに起因する差別的取扱いにつき、米国では保険会社を被告とした訴訟が提起されている。争点の類型として、①保険金支払段階における不正検知スコア等のアルゴリズム的意思決定ツールが黒人契約者の請求を過剰に疑義扱いし遅延させるとして連邦公正住宅法 (FHA) 上の disparate impact が主張された事例 (Huskey v. State Farm Fire & Casualty Co., No. 22 C 7014 (N.D. Ill. filed Dec. 14, 2022)) と、②引受・料率段階における信用情報ベースの自動スコアリングがマイノリティに不均衡な高保険料等をもたらすとして争われた事例 (DeHoyos v. Allstate Ins. Co., No. SA-01-CA-1010-FB (W.D. Te

外部起因のリスクとして、AI を悪用したサイバー攻撃の高度化がある¹⁰。生成 AI により巧妙なフィッシングメールやマルウェアが作成可能となり、サイバー攻撃の脅威は増している。

以上のリスクに横断的に関わる構造的問題がブラックボックス性である¹¹。深層学習モデルの判断過程は人間には理解困難であり、保険事故発生時の因果関係立証において重大な障害となりかねない。集積リスクとして、少数の基盤モデルへの広範な依存はシステミック・リスクを内包し、特定モデルの欠陥が多数の企業に同時損害を引き起こしかねない¹²。開発から運用に至るまでの全プロセスを通じ、学習データや推論結果に含まれる機密情報の意図せぬ流出リスクも存する¹³。

2 従来型保険の限界

従来、企業はデジタルリスクへの備えとしてサイバー保険を活用してきた。サイバー保険は、主に不正アクセスやマルウェア感染といった外部からのサイバー攻撃による損害を填補する。しかし、上述のような AI 特有のリスク、とりわけ AI の自律的判断や出力内容の瑕疵に起因する損害は、従来のサイバー保険や賠償責任保険でカバーするのが困難なケースが多い。だからこそサイバー保険の発展過程が参照に値する。サイバー保険は AI 保険に対する補償の前身であるだけでなく、リスクの構造において AI リスクと類似性を有する。つまり、不確実性の高い新技術リスクである点、システミックリスクの可能性を有する点、外部攻撃、内部の瑕疵や利用者の行為の複合によって損害が生じる点で、両者は共通の構造を

x. 2001) (和解で終結) ; Ojo v. Farmers Group, Inc., No. 06-55522 (9th Cir.) (Ojo, et al. v. Farmers Group, Inc., 356 S.W.3d 421 (Tex. 2011)参照) がある。

¹⁰ 濱田和博「欧米等におけるサイバーリスク・保険の現状・課題および保険業界の取組み」損保総研レポート150号(2025年)13-14頁参照。同論文は、ニューヨーク州金融サービス局(DFS)のガイダンスに基づき、AIを利用したソーシャルエンジニアリングの高度化、サイバーセキュリティ攻撃の有効性・規模・速度の増幅、非公開情報の大量漏洩、サプライチェーン依存による脆弱性の増大を主要リスクとして整理する。

¹¹ ブラックボックス問題とは、深層学習等の高度な AI モデルにおいて、入力から出力に至る判断過程が人間にとって理解困難である状態をいう。

¹² システミック・リスクとは、特定の基盤モデルの欠陥が、当該モデルを利用する多数の企業において同時に損害を引き起こす可能性をいう。

¹³ AI 開発・運用過程においては、学習データや推論結果に含まれる機密情報が意図せず外部に漏洩するリスクがある。

持つといえる。また後述するように海外の保険実務上、現在の AI 除外条項はサイバー保険約款の改訂過程でも出現しており、AI 保険はサイバー保険の後継として市場で位置づけられている。PL 保険や CGL 保険等も約款紛争を経て精緻化されてきたが、サイバーと AI の間にはこうした構造的・実務的な連続性があり、他の保険種目には見られない固有の参照可能性が認められる。では、サイバー保険はいかなる発展を遂げ、AI 保険の設計にいかなる教訓を残したか¹⁴。

III サイバー保険からの教訓

海外ではサイバー保険は約款の曖昧さから多くの紛争を経験し、裁判例を通じて解釈が精緻化されてきた。無論、新技術リスクに起因する保険約款の精緻化という現象はサイバー保険に限られたものではなく、PL 保険や CGL 保険においても同様の過程が見られる。しかし前章で述べたとおり、サイバー保険はリスク構造の類似性および実務上の連続性において AI 保険と特別な関係にあると評価することができ、その発展過程から得られる教訓は AI 保険の設計に応用しうる。保険による規律付けが機能するには、まず約款が AI に関する実態に即して精緻化されている必要がある。サイバー保険の発展過程はその精緻化がいかに進むかを示す先例であると考ええる。

1 サイバー保険市場の概況

サイバー保険市場は、サイバー攻撃の増加と巧妙化に伴い急速に変化している。米国等では加入率が上昇する一方、損害率の上昇により保険料率の引上げや引受条件の厳格化が進んでいる¹⁵。

¹⁴ なお、本稿は主として第三者に対する賠償責任を念頭に置いているが、AI システム自体の瑕疵・汚染に対する first party coverage も問題となりうる。学習データの汚染によって生じた欠陥モデルの廃棄・再学習費用、汚染済みモデルが出力した情報の回収・訂正費用等は従来の賠償責任保険や事業中断保険においても必ずしもカバーされない空白領域を形成する可能性がある。サイバー保険においては既にデータ復旧費用・システム復旧費用条項が普及しており、その延長として AI 固有のモデル復旧費用を補償対象とする条項設計の可能性を検討する余地がある。

¹⁵ サイバー保険市場においては、2021 年から 2022 年にかけて保険料率の引上げ傾向が顕著であった (NAIC, Report on the Cybersecurity Insurance Market (Nov. 3, 2023) 7 頁)。同報告書によれば、米国元受保険市場における保険料は 2021 年に約 75%、2022 年に約 50% 増加した。

2 戦争免責条項の再定義 Merck 判決

Merck & Co. v. Ace American Insurance Co. 判決は、NotPetya サイバー攻撃による損害について戦争免責条項の適用が争われた事案である¹⁶。裁判所は、戦争の概念をサイバー空間に無限定に拡大すべきではないとして、原告 Merck の請求を認容した。もっとも、本件は 2024 年 1 月にニュージャージー州最高裁での口頭弁論直前に和解で終結しており、最高裁による最終的な法理の確定は回避された。控訴審判決たる戦争免責不適用の先例的価値は残るものの、保険会社側が不利な先例の確定を回避したとみることができる。戦争やテロといった伝統的な免責概念をサイバー空間にどう適用するかという根本問題が提起され、その後の約款改定を促す契機となったことは否定できない。Merck 事件をはじめとする国家関与型サイバー攻撃をめぐる市場紛争を背景として、ロイズ・マーケット協会 (LMA) は国家関与型サイバー攻撃を明示的に免責とするモデル条項を策定していた¹⁷。曖昧な概念に依存せず、免責範囲を具体的に規定する方向への転換である。

3 コンピュータ詐欺条項の射程 G&G Oil 判決

G&G Oil Co. of Indiana, Inc. v. Continental Western Insurance Co. 判決では、ランサムウェア攻撃による身代金支払いが、商業犯罪保険のコンピュータ詐欺条項におけるコンピュータの使用により詐欺的に金銭の移転を引き起こしたことに直接起因する損失に該当するか、が争われた¹⁸。インディアナ州最高裁は、「詐欺的に移転を引き起こす」を「策略により取得すること」と広く解し得るとした上で、身代金支払いに至る過程で FBI への相談等の介在行為があったとしても因果的近接性は維持されるとして、保険会社勝訴の原審を破棄し差し戻した。本判決は、サイバー保険ではない従来型の商業犯罪保険がランサムウェア損害を

¹⁶ Merck & Co., Inc. v. Ace American Insurance Co., Nos. A-1879-21, A-1882-21 (N.J. Super. Ct. App. Div. May 1, 2023). NotPetya マルウェア攻撃 (2017 年) による約 14 億ドルの損害について、全危険型財産保険の戦争免責条項の適用が争われた事案である。ニュージャージー州控訴審は Merck の請求を認容したが、州最高裁での口頭弁論直前の 2024 年 1 月に和解で終結した (和解条件は非公表)。

¹⁷ 濱田・前掲注 4) 14-18 参照。LMA は国家関与型サイバー攻撃を明示的に免責とするモデル条項案を公表している。

¹⁸ G&G Oil Co. of Indiana, Inc. v. Continental Western Insurance Co., No. 20S-PL-617, 2021 Ind. LEXIS 182 (Ind. Mar. 18, 2021). 本件は商業犯罪保険 (Commercial Crime Coverage) のコンピュータ詐欺条項の解釈が争われた事案であり、争点は「fraudulently cause a transfer」および「resulting directly」の意義である。

カバーし得ることを示した点で、サイレント・サイバー問題の典型例であると評価できる。

4 小括

これらの判例は、従来の保険約款が現代の複雑なデジタルリスクに対応しきれないことを示している。Merck 事件は戦争免責条項のサイバー空間への適用という約款解釈上の問題を提起した。また、G&G Oil 判決は従来型の商業犯罪保険がランサムウェア損害をカバーし得ることを示すことでサイレント・サイバー問題の深刻さを顕在化させた。いずれも、具体的で限定的な約款設計への移行を促した点で共通する。AI 保険でも、AI 起因の意味を明確に定義し、免責範囲を具体化する必要がある。リスクの具体化と約款の精緻化というプロセスは、AI 保険もサイバー保険と同様の道をたどることになると考える。

IV 海外における AI 保険の動向

約款の精緻化という課題を踏まえつつ、現在の海外市場は AI リスクをいかに扱おうとしているか。その動向は保険による規律付けの実現可能性を測ることに大いに役立つ。そして、AI リスクの不確実性を前に海外の保険市場は除外と専用化という二方向へ分化しつつある。

1 AI 除外条項とサイレント AI 問題

一方の潮流は、既存保険からの AI リスク排除である。米国では複数の保険大手グループが、AI ツールを導入する企業関連で生じる賠償責任を免責とする許可を規制当局に申請している¹⁹。急速に発展する AI 技術が将来引き起こしかねない巨額損害への警戒がその背景にある。Bennett & Corson は、更改シーズンにおいてサイバー保険やメディア賠償責任保険に限らず広汎な AI 除外条項が各種保険ラインに入り込むことへの警戒を説いている²⁰。

¹⁹ 米国では、複数の保険大手グループが、AI ツールを導入する企業に関連して生じる賠償責任を補償対象外とする許可を規制当局に申請している。(Lee Harris & Cristina Criddle, Insurers Retreat from AI Cover as Risk of Multibillion-dollar Claims Mounts, FIN. TIMES, Nov. 23, 2025, <https://www.ft.com/content/abfe9741-f438-4ed6-a673-075ec177dc62>.)

²⁰ Lynda Bennett=Alexander Corson 「Policyholders Should Negotiate to Limit AI Exclusions at Renewal」 Bloomberg Law (2025 年 12 月 15 日) (<https://news.bloomberglaw.com/legal-exchange-insights-and-commentary/policyholders-should-negotiate-to-limit-ai-ex>)

AI 除外条項の問題の本質は、AI の輪郭が曖昧でかつ広範な範囲に遍在している点にある。具体的に、初期の AI 除外条項は、定義が曖昧であること、原因関係の入口文言が広いこと、過剰に包括的であること、という特徴を持つ。²¹企業活動において AI が関与しない業務の方が希少になりつつある以上、これらの特徴は深刻な問題を引き起こす可能性を秘めている。除外範囲の過大さが問題の核心である²²。具体的には、AI のあらゆる使用に関わる請求を排除する条項案により、AI が周縁的に使用されているに過ぎない事故まで免責となりかねない。例えば、専門職賠償に由来する請求であっても、保険契約者が AI でマーケティング資料を作成していたという周辺事情を根拠に免責されるリスクがある。AI が業務プロセス全般に浸透した場合には、事実上あらゆる請求が免責対象となりうる²³。

この構造は、サイバー保険市場において先行して顕在化したサイレント・サイバー問題と同様である²⁴。サイバーリスクが従来型保険で意図せず補償される、あるいは補償されないという問題が生じたように、現在はサイレント AI リスクの排除が進行している。ここでサイレント AI とは、約款の曖昧さにより、保険会社が意

clusions-at-renewal, 2026 年 3 月 12 日最終閲覧)。同論考は、Lowenstein Sandler 法律事務所
の保険回収専門によるものである。

²¹ Bennett=Corson・前掲注 20) 参照。

²² Bennett=Corson・前掲注 20) 参照。

²³ AI 除外条項は、サイバー保険やメディア賠償責任保険に限られず、D&O 保険や E&O 保険等
にも波及し得る。Bennett=Corson・前掲注 20) 参照。

²⁴ サイレント・サイバーとは、サイバーリスクを明示的に補償対象としていない保険契約に
おいて、サイバー関連損害が意図せず補償される、または補償されない状態をいう。この問題
は 2010 年代半ばから英国ロイズ市場において議論が本格化し、ロイズは 2019 年 7 月にサイレ
ント・サイバーリスクの特定と対応を市場参加者に求めるガイダンスを公表している (Lloyd'
s, Market Bulletin Y5258, Providing clarity for Lloyd's customers on coverage for cy
ber exposures (4 July 2019)。同文書の要求事項はその後継文書である Lloyd's, Market Bu
lletin Y5277, Update - Providing clarity for Lloyd's customers on coverage for cyber
exposures (29 January 2020), at 1 においても確認される。(https://assets.lloyds.com
/assets/y5277-update-providing-clarity-for-lloyds-customers-on-coverage-for-cyber-ex
posures/1/Y5277%20Update%20%20Providing%20clarity%20for%20Lloyds%20customers%20on%20
coverage%20for%20cyber%20exposures.pdf (2026 年 2 月 24 日最終閲覧))。邦語文献とし
て、金・前掲注 4) 参照。

図せず AI リスクを引受けてしまっている、あるいはその逆の状態を指す²⁵。広範な AI 除外条項の拡大は保険の空白を生じさせる恐れがある²⁶。

2 AI 専用保険の開発とパラメトリック的設計

もう一方の潮流は、AI リスクを肯定的に引き受ける専用保険の開発である。前章で検討した約款精緻化という観点で括れば方向性は同じである。

カナダの Armilla AI とロイズ・オブ・ロンドンの Chaucer が共同開発した Affirmative AI Liability Insurance は、ハルシネーションやモデルドリフト等による損失を明示的に補償対象とする²⁷。補償トリガーの設計が特徴的であり、当該保険は、AI ツールの精度が事前設定の基準を下回った場合や、期待性能の未達により損失が発生した場合を補償発動要件とする²⁸。従来の損害発生とは異なるパラメトリック保険的な設計であり、第 V 章で後述する因果関係論の限界を克服する試みにほかならない。Armilla AI の商品構造は保険と保証の二層からなる²⁹。同社は AI 導入企業向けの Armilla Insured (=賠償責任保険) とは別に、AI ベンダー向けの Armilla Guaranteed (=パフォーマンス保証) を提供する。後者は AI が契約上の KPI を満たさない場合に買い手への返金を保証するものであるが、同社は「保証は契約上の履行保証であり、保険契約ではない」と明言する。しかし、この保証の履行能力は Swiss Re、Greenlight Re、Chaucer 等の保険会社・再保険会社がバックアップしている³⁰。

²⁵ これは約款の問題であり、後述する企業が管理していない従業員の勝手な AI 利用を指すシャドーAI とは区別される。

²⁶ ここでは保険の空白とは、既存の保険契約ではカバーされない領域を指す。

²⁷ Armilla AI とロイズ・オブ・ロンドンの Chaucer が共同開発した「Affirmative AI Liability Insurance」。

²⁸ 同保険では、AI ツールの精度が事前に設定された基準を下回った場合や、期待性能の未達によって損失が発生した場合を補償トリガーとしている。

²⁹ Armilla AI, “AI Insurance - Lloyd’s Coverholder”, available at <https://www.armilla.ai/ai-insurance> (2026 年 3 月 12 日最終閲覧)。同ウェブサイトの FAQ では、“Is this Warranty an Insurance Product? No. A warranty is a performance guarantee made in contract, not an insurance policy.” と明記されている。

³⁰ 保険でバックアップされた保証という構造それ自体は珍しくない。日本においても、家電量販店が消費者に延長保証を有償で提供しつつ、そのリスクを損害保険会社に移転する形態が広く普及しており、家電量販店と消費者との契約関係という保証と、家電量販店と保険会社との契約関係という保険の組み合わせは実際に確立している。Armilla Guaranteed について日本法上問題となるのは、この構造の一般的性格ではなく、リスクの最終的な引受主体が日本に

AI 専用保険が登場した背景には、ハルシネーションがもたらす事業リスクの顕在化がある。エア・カナダ社では、AI チャットボットが遺族向け割引運賃について旅行完了後にも遡及的に申請可能である旨の誤った案内を行い、顧客がその案内に基づいて航空券を購入した³¹。同社は、チャットボットは別の法的主体でありその発言について責任を負わないと主張したが、ブリティッシュコロンビア州民事紛争解決審判所（BCCRT）はこれを退け、企業がウェブサイト上の全ての情報について責任を負うと決定した。本件は BCCRT の決定であり法的拘束力は限定的であるものの、AI エージェントの行為を企業の責任に帰属させた象徴的な事例として実務に衝撃を与えた。第VI章で検討する Agentic AI の責任帰属問題にも示唆を与える。同様の問題は英国の Virgin Money や DPD でも発生しており、チャットボットの不適切対応が企業の信用問題に発展している³²³³³⁴。

において保険業の免許を有しない海外再保険会社であるという点にある。保険業法 185 条 1 項は外国保険業者が日本で保険業を営むには日本に支店等を設けて内閣総理大臣の免許を受けることを要求し、同法 186 条 1 項は免許を有しない外国保険業者による海外直接付保を原則として禁止する。Armillia Guaranteed においては、形式上の保証提供者である Armilla AI 自身はリスクを最終的に負担せず、Swiss Re・Greenlight Re・Chaucer 等の海外再保険会社がバックストップを提供している。これらの海外再保険会社が、日本に所在する企業との間で実質的なリスク引受けを行う場合、保険業法 185 条 1 項の免許要件および同法 186 条 1 項の海外直接付保禁止規定との関係が問題となりうる。もっとも、当該取引が専ら海外で完結する場合や、再保険取引として構成される場合には規制の射程が異なる可能性があるため、慎重な検討を要する。なお、保険でバックアップされた保証が同法 2 条 1 項所定の「保険業」に該当するか否かは、リスク移転の実質に照らして判断されるべき別論点であり、本稿では検討を留保する。

³¹ Moffatt v. Air Canada, 2024 BCCRT 149.

³² Financial Times, “Insurers launch cover for losses caused by AI chatbot errors” (May 11, 2025), available at <https://www.ft.com/content/1d35759f-f2a9-46c4-904b-4a78ccc027df> (2026 年 3 月 12 日最終閲覧)。Tom Graham 氏 (Chaucer パートナーシップ担当責任者) は、故障に過度に陥りやすい (excessively prone to breakdown) AI システムには引受けを行わないとした上で、“We will be selective, like any other insurance company”と述べている。

³³ Fortune, “Virgin Money chatbot scolds customer who ‘confused bank’s name for insult’” (Jan. 30, 2025), available at <https://fortune.com/europe/2025/01/30/virgin-money-chatbot-scolds-customer-confuse-banks-name-insult/> (2026 年 3 月 12 日閲覧)。

³⁴ The Guardian, “DPD AI chatbot swears, calls itself ‘useless’ and criticises firm” (Jan. 20, 2024), available at <https://www.theguardian.com/technology/2024/jan/20/dpd-ai-chatbot-swears-calls-itself-useless-and-criticises-firm> (2026 年 3 月 12 日閲覧)。

Armilla AI と Chaucer の保険商品は、AI リスクを事前に評価・管理しつつ損害発生時に補償を提供する。Chaucer のパートナーシップ担当責任者である Tom Graham 氏は、故障に過度に陥りやすいと判断される AI システムについては引受けを行わない方針を示し、“We will be selective, like any other insurance company”と述べている³⁵。被保険者が用いる AI システムが一定の評価基準を満たすことが引受の条件であり、不安定または不正確なモデルには保険が適用されない。保険会社が AI ガバナンスのゲートキーパーとなっており、第VI章で検討する保険による規律付けの具体例であるといえる。2026年1月の改訂で補償範囲が拡張され、最大2,500万ドルという高額な保険限度額での提供も可能となった³⁶。改訂後の約款では、AI モデルのハルシネーション・不正確さ・性能劣化による第三者の経済的損失、名誉毀損・営業秘密露出・守秘義務違反を含む有害出力に起因する請求、自律 AI エージェントの誤判断や不適切なツール使用またはエスカレーション・エラーによる請求、AI の出力を通じた意図しない情報漏えいに起因するプライバシー事故、AI を搭載した自動化システムによる物的損害、EU AI Act 等の AI 規制法に基づく当局調査への対応費用および填補可能な行政制裁金が明示的に補償対象となった³⁷。自律 AI エージェントに関する補償は第VI章6節で検討する Agentic AI の責任問題に対応し、AI 規制法違反への対応費用の補償は AI 規制の国際的な進展を見据えた設計である。契約形態はクレームズ・メイド方式で、保険期間中に通知されたクレームに補償が提供される。Armilla AI は、サイバー保険・E&O 保険・一般賠償責任保険の約款を流用するのではなく、AI システムの失敗メカニズムに沿ってゼロから設計したと説明しており³⁸、第III章で検討した約款精緻化の方向を体現している。

ミュンヘン再保険 (Munich Re) では、技術デューデリジェンスを通じて AI モデルの品質を評価し、その信頼性を保険で裏書きするパフォーマンス保証保険を提

³⁵ Financial Times・前掲注 32) 参照。

³⁶ Armilla AI, “Armilla AI Raises Lloyd’s-Backed Coverage to \$25M as Traditional Insurers Retreat from AI Risk” (Jan. 21, 2026), available at <https://www.armilla.ai/resources/armilla-ai-raises-lloyds-backed-coverage-to-25m-as-traditional-insurers-retreat-from-ai-risk> (2026年3月12日最終閲覧)。

³⁷ 改訂後の約款は、従来のエンドースメント型から脱却した包括的な積極補償 (affirmative coverage) の枠組みに刷新されている。

³⁸ Armilla AI, Press Release “Armilla Launches Affirmative AI Liability Insurance with Lloyd’s Underwriter, Chaucer” (Apr. 30, 2025), available at <https://www.armilla.ai/resources/armilla-launches-affirmative-ai-liability-insurance-with-lloyds-underwriter-chaucer> (2026年3月12日閲覧)。同社 CEO の Karthik Ramakrishnan は、“It’s built from the ground up to address the specific ways AI can fail”と述べている。

供している³⁹。同社のデューデリジェンスは導入ミーティング、詳細技術質問票、フォローアップ・ミーティング、報告・結論の4段階で構成され、クライアントと合意したKPIが補償発動の基準となる⁴⁰。同社によれば、技術デューデリジェンスを受けた企業の90%が保険契約の提供を受けている。同社はパフォーマンス保証保険を信頼の加速装置と位置づける⁴¹。保険会社がAIモデルの品質を裏書きすることでAI導入への顧客の信頼を高め、イノベーション普及を促す。この保険による信頼構築は、第VI章で検討する保険による規律付けと関連する⁴²。

3 小括

これら二つの潮流は同じ保険市場のメカニズムの表裏の関係であるといえる。従来型保険ではAIリスクは不確実性が高く、サイレントAI回避のため除外せざるを得ない。他方、デューデリジェンスを経て切り出されたリスクは専用保険としてプライシングが可能となる。現在の海外動向は、AIリスクが漠然とカバーされる段階から明示的に管理・移転される段階への過渡期に位置する。この二極化は同時に、リスク評価の困難、因果関係立証の複雑さ、保険と企業ガバナンスの関係といったAI保険固有の課題を顕在化させる。

V AI保険の確立に向けた課題

前述の通り、AI保険のニーズは高まっているものの、市場は黎明期にある。普及にあたっては、従来のシステムやサイバーリスク保険とは異なるAI特有の課題の克服が前提条件となる。海外の動向が示すように、AI保険の普及には技術的・法的課題の克服が前提となる。これらの課題が解決されなければ、保険による規律付けは実現困難である。

³⁹ Munich Re・前掲注1) 参照。

⁴⁰ Munich Re・前掲注1) 4-5頁。

⁴¹ Munich Re・前掲注1) 7頁。

⁴² 日本ではあいおいニッセイ同和損保が2024年2月に国内初の生成AI専用保険を公表した(あいおいニッセイ同和損害保険株式会社「【国内初】生成AIのリスクを補償する『生成AI専用保険』の提供開始」(2024年2月28日)) https://www.aioinissaydowa.co.jp/corporate/about/news/pdf/2024/news_2024022701277.pdf (2026年3月12日最終閲覧)。生成AI利用に伴う知財侵害、情報漏えい、ハルシネーション関連の対外対応費用等を補償対象とする。もっとも、同保険はAI開発会社であるArchaic社を保険契約者とし、同社が開発した生成AIシステムを利用する企業のみを被保険者とする仕組みであり、一般企業が広く加入できる汎用商品ではない。

1 リスク評価と因果関係立証の困難

そもそも、リスクの定量的評価が難しい。AIの挙動は確率論的でブラックボックス性が高く、事故発生確率の予測が困難である。十分な損害データの蓄積がない現状では、大数の法則に基づく伝統的な保険料の算出モデルは機能しにくい。また、因果関係の立証も複雑である。AIによる損害が学習データのバイアスによるのか、ファインチューニングの不備によるのか、ユーザーの悪意あるプロンプト注入によるのかを区別することは技術的に容易ではない。加えて、責任の所在も不明確になりやすい。従来の技術面のエラーであれば責任主体を容易に特定できたが、AIリスクには開発者、ベンダーやユーザーなど多数の関係者が関与するため、責任の特定が格段に困難となる。

そして、保険引受には高度な技術デューデリジェンスが必要となる。保険会社には従来のリスク評価手法に加え、情報工学の専門知見が求められることになる。前章で検討したMunich Reの技術デューデリジェンスプロセスが一つの海外における先行モデルとなる。

2 間接差別と代替変数の問題

保険業界特有の課題として、AIによる間接差別がある⁴³。AIが膨大な非構造化データからリスクを推論する際、人種、信条、社会的身分といった法的に考慮してはならない属性と相関する代替変数を用いて差別的選別を行う危険がある。企業はそれを事前に検知し排除する体制を構築せねばならない⁴⁴。この課題は第VI章で検討する内部統制システム構築義務とAIガバナンスの問題と関連する。

3 因果関係論の限界とパラメトリック保険

より根本的な問題として、従来の保険法理が前提としてきた因果関係論は、AIリスクにおいて機能し得るのか。

AIリスクが発現する場面は多岐にわたる⁴⁵。賠償責任のみならず、ビジネスチャンス消失、第三者の名誉毀損等の間接損害も含まれうる。AIシステムが引き起こ

⁴³ AIが膨大な非構造化データからリスクを推論する際、法的に考慮してはならない属性と相関する代替変数を用いて、差別的な選別を行うリスクがある点については、前掲注9参照。

⁴⁴ 代替変数による間接差別を防止するには、AIモデルの出力が保護属性と相関していないかを検証するバイアス監査の仕組みを構築することが求められる。

⁴⁵ AIリスクの発現形態は多岐にわたる。賠償責任のみならず、ビジネスチャンス消失、第三者の名誉棄損等の間接損害も含まれ得る。

す損害と AI システムの挙動との間に、従来想定されていたような因果関係を立証することは、ブラックボックス性ゆえに困難な場合が多い。全ての損害について因果関係の立証を求めれば、被保険者保護や保険金支払の迅速性が損なわれる。

日本の保険法の強行法規はこの問題を深刻化させる。保険法 31 条 2 項 1 号但書は、告知義務違反があっても当該事実と保険事故との間に因果関係がなければ保険者は免責されないとする因果関係不存在特則を定める。この特則は片面的強行規定であり（保険法 36 条）、被保険者に不利な方向、すなわち保険者の解除権を拡張する方向での約款による変更は認められない。他方、被保険者に有利な方向での変更は許容される。もっとも、AI 保険は事業者向け取引を主たる対象とするところ、保険法 36 条 4 号は事業者間の保険契約について同法の強行規定の適用を除外する余地を認めており、当事者自治による契約設計の可能性も検討に値する。この強行規定と AI のブラックボックス性は、被保険者保護の観点から問題を生じさせる。因果関係不存在特則の立証責任は被保険者側にある。すなわち、保険者が告知義務違反を理由として契約を解除した場合、被保険者が告知しなかった事実と保険事故との間に因果関係が存在しないことを積極的に立証しなければ、同特則によって保護されない。AI のブラックボックス性ゆえに、被保険者がこの立証に成功することは技術的に困難である。例えば、基盤モデルのバージョンについて不告知があった場合、被保険者が当該バージョンの相違と今回の事故との間に因果関係がないことを論証するには、モデルの内部動作を解析する必要があるが、深層学習モデルにおいてその解析は現状では実現困難である。その結果、因果関係不存在特則は機能しづらく、保険者の解除権行使が容易になる反面、告知義務の射程が技術的知見の乏しい被保険者には予測困難な AI 関連事項にまで及ぶことで、被保険者保護に空白が生じる可能性がある。

この強行規定の問題を法的に回避する手段として、パラメトリック保険が考えられる⁴⁶。パラメトリック保険は、実際の損害額ではなく事前に定めた指標が一定の閾値を超えた場合にあらかじめ合意された金額を支払う保険形態である。AI リスクでは、モデル精度が一定%を下回った場合や、システムダウンタイムが一定時間を超えた場合等の客観的指標に基づく設計が考えられる。前章で検討した Armilla AI と Chaucer の Affirmative AI Liability Insurance が採用した性能低下や期待

⁴⁶ パラメトリック保険とは、実際の損害額ではなく、事前に定めた指標が一定の閾値を超えた場合に支払う保険形態である（吉澤卓哉「インデックス保険の「保険」該当性——定額給付型の損害保険契約——」産大法学 53 卷 3・4 号 187 頁、207 頁注 39（2020 年）（パラメトリック保険はインデックス保険と同義に用いられるとする）。インデックス保険ないしパラメトリック保険の概念については、吉澤卓哉「情報社会の急速な進展による保険制度における「信頼」の変容」保険学雑誌 649 号 173 頁、181 頁（2020 年）も参照。）。

性能の未達という補償トリガーは、このパラメトリック的発想の表れであると評価できる。

もつとも、パラメトリック保険の保険法上の位置づけには検討を要する。保険法2条6号は損害保険契約を「一定の偶然の事故によって生ずることのある損害をてん補することを約する」契約と定義する。パラメトリック保険では、パラメータの発動と実損害との間に乖離が生じ得るため、損害てん補性との関係が問題となる。この問題につき、吉澤卓哉の定額給付型損害保険契約に関する分析が示唆に富む⁴⁷。吉澤は、損害保険契約の保険給付要件を、被保険利益の存在（要件A）、保険事故の発生（要件B）、損害の発生（要件C）、損害額を超えない保険給付（要件D）の四要件に整理したうえで、要件Cおよび要件Dのそれぞれについて厳格な利得禁止原則と緩やかな利得禁止原則を区別する。すなわち、要件Cについては、現実の損害発生を要件とするもの（C1）と、保険事故の発生をもって損害が発生したとみなすもの（C2：損害のみなし発生）に、要件Dについては、具体的に損害額を算定するもの（D1）と、少なくとも一定額の損害が発生したとみなすもの（D2：損害額のみなし算定）に分類する。そして、利得禁止原則の目的である賭博禁止とモラル・ハザード防止の双方が確保される限り、緩やかな利得禁止原則（C2・D2）の適用が認められ、定額給付型の損害保険契約も有効であるとする。吉澤は、当該枠組みに基づき、農業分野のインデックス保険、航空機遅延保険、自然災害に関するインデックス保険等の実在する保険商品を検討し、いずれも損害保険契約としての法的性質を有すると結論づけている。

この分析枠組みをAI保険に適用すると、以下の整理が可能である。前章で検討した Armilla AI/Chaucer の Affirmative AI Liability Insurance が採用する性能低下トリガーは、吉澤のいうⅢ類型である損害のみなし発生および損害額のみなし算定を行う損害保険契約に位置づけられる⁴⁸。つまり、企業がAIシステムの

⁴⁷ 吉澤卓哉「インデックス保険の『保険』該当性——定額給付型の損害保険契約——」産大法学53巻3・4号（2020年）187頁以下。同論文は、定額給付型の損害保険契約がどこまで認められるかを、利得禁止原則の目的である賭博禁止とモラル・ハザード防止の観点から体系的に検討する。なお、山下友信は、モラル・ハザードが生じ得ない取引であれば利得禁止原則を適用する必要はないという帰結を導き得ると述べる（山下友信「保険・保険デリバティブ・賭博 リスク移転取引のボーダー」江頭憲治郎＝増井良啓編『融ける境 超える法3 市場と組織』（東京大学出版会、2005年）243-244頁）。吉澤はこの山下説を踏まえつつ、利得禁止原則を廃棄するのではなく緩やかな利得禁止原則として維持する立場を採る。

⁴⁸ 吉澤・前掲注47）213-214頁。吉澤は、Ⅲ類型の損害保険契約について、保険事故が発生した場合にほとんどの被保険者に損害が発生しており、通常発生する損害額以下の定額給付を受けたとしても利得は生じないことから、賭博保険となる惧れは低いとする。農業分野のインデックス保険については、Ⅰ類型の典型的な保険商品におけるモラル・ハザードの問題がむしろ解消されたと評価する（同214頁）。

運用について被保険利益を有し（要件 A）、AI モデルの精度が事前設定の基準を下回るといふ保険事故が発生した場合に（要件 B）、現実の損害発生の確認を経ずに損害の発生をみなし（要件 C2）、具体的な損害額の算定を行わずにあらかじめ合意された金額を保険金として支払う（要件 D2）という構造である。

賭博禁止の観点からは、被保険利益（要件 A）および保険事故の発生（要件 B）の充足が求められており、かつ、AI モデルの精度が事前設定の基準を下回った場合に企業に一定の事業損害が発生する蓋然性は高い。定額で支払われる保険金の額が保険事故によって通常発生する損害額を超えない限り、賭博保険となるおそれは低い。モラル・ハザード防止の観点からは、吉澤が指摘するとおり、定額給付型の損害保険契約では事故便乗型のモラル・ハザードが構造的に発生しない点に注目すべきである。さらに、AI モデルの精度低下は、自然災害に類似して被保険者が容易に操作し得ない客観的指標であり、事故偽装型のモラル・ハザードも想定しにくい。もちろん、被保険者が意図的にモデルの品質を劣化させる可能性は皆無ではない。しかし、企業にとって自社の事業基盤である AI モデルを意図的に劣化させることは、顧客からの信用喪失や事業上のビジネス機会の損失を招くものであり、定額保険金の取得と比較して経済合理性を欠く。加えて、第 IV 章で検討した Munich Re の技術デューデリジェンスや Armilla AI の引受審査が、このリスクを外的に統制している。

吉澤の分析枠組みに依拠すれば、パラメトリック型の AI 保険は、被保険利益が存在し、パラメータと損害との間に合理的な相関関係が認められ、かつ、賭博禁止とモラル・ハザード防止が確保される限り、損害保険契約としての性質を維持すると解される。ただし、パラメータの設定が損害との相関を欠く場合には、もはや利得禁止原則の緩和では正当化が困難であり、保険契約ではなくデリバティブ取引として規制される可能性がある点には留意する必要がある⁴⁹⁵⁰。

⁴⁹ AI リスクの継続的・反復的発生という性質は、総支払限度額および填補限度額の自動復元条項の設計の要請という課題を提起する。ハルシネーションやモデルドリフトに起因する事故は保険期間内に反復的に発生する可能性があり、総支払限度額の設定および填補限度額の自動復元の要否・内容を約款上明示することが保険契約者保護の観点から求められる。本稿は単一の保険事故を念頭に置く議論を中心とするが、反復的損害への対応については今後の検討課題としたい。

⁵⁰ 吉澤・前掲注 47) 222 頁は、Ⅲ類型のインデックス保険であっても、緩やかな利得禁止原則が認められる一定の条件を充たさないものは損害保険契約とは認められないとする。また、山下友信『保険法（上）』（有斐閣、2018 年）29 頁は、保険と保険デリバティブとの区別について、損害をてん補するものか損害の発生の有無を問わない金銭の支払をするものかで区別されるとする。吉澤はこの区別を前提としつつ、被保険利益が存在し保険事故と損害との間に合理的相関がある限り、定額給付であっても損害をてん補の範疇にとどまると解する。

4 小括

AI 保険の確立には、リスク評価手法の開発、因果関係論の再構成やパラメトリック保険の活用等が必要である。しかし、技術的・法的課題の解決だけでは足りない。AI 保険が実効的に機能するには企業における AI ガバナンス体制の構築が前提となり、保険はそのガバナンス構築を促す機能を担う。第IV章で検討した Munich Re や Armilla AI/Chaucer が引受条件として AI ガバナンス体制を求めているのは、まさにこの認識に基づくものと評価できる。では、AI ガバナンス体制とは法的にいかなる内容を持ち、取締役はいかなる義務を負うか。以上の検討を踏まえ、保険による規律付けとしての AI 保険が機能するための条件の整理を試みる。

VI 日本法への含意

1 告知義務と AI インベントリ

保険法 4 条は危険に関する重要な事項について告知義務を課している⁵¹。第 II 章で整理した AI リスクの特性を踏まえれば、生成 AI の利用有無、利用目的、AI ガバナンス構築状況が危険評価の重要な要素となる。どのような基盤モデルを採用しているか、RAG によるハルシネーション防止策を講じているか⁵²、AI モデルの監視・更新体制が整備されているか等の技術的な仕様が告知事項として問われる可能性がある。もっとも、それにあたり告知義務の法的構造を把握する必要がある。周知の通り保険法の制定により、告知義務は自発的申告義務から質問応答義務へ転換した。この転換の帰結として、保険者が質問しなかった事項については原則として告知義務違反を問えない⁵³。そして、その他重要な事項等の抽象的・包括的な質問の効力も限定的に解される⁵⁴。

これを AI 保険にあてはめると、保険者が AI の利用の有無を問うのみで、基盤モデルの種類、RAG の利用有無、ファインチューニングの有無等の技術的仕様について具体的で明瞭な質問を設定しない限り、企業のシャドーAI 利用や不適切なモデル選定は告知義務違反を構成しない可能性が高い。AI インベントリの整備は企業

⁵¹ 保険法 4 条は、保険契約者等に対し、危険に関する重要な事項について告知義務を課している。

⁵² RAG とは、外部データベースから関連情報を検索し、それを生成 AI の出力に組み込むことで、ハルシネーションを抑制する技術をいう。

⁵³ 山下友信『保険法（上）』（有斐閣、2018 年）397 頁。

⁵⁴ 山下・前掲注 53) 412 頁。

の役割であり義務ではあるものの、保険者が告知義務違反に基づく解除という効果を発生させるためには、保険者側に AI に関するリスクを具体的に問う力、技術的知見に基づく質問を設定する能力が求められる。第IV章で検討した Munich Re の技術デューデリジェンスは、保険者の問う責任を制度化したものと解することができる。

企業には、正式承認なく従業員が使用するシャドーAI⁵⁵を含め、AI 利用実態を把握する AI インベントリ管理が法務リスク管理として求められる。ここでの棚卸しは単なる社内コンプライアンスのためのものではなく、保険の観点では、質問票における AI 利用についての回答の不正確さが、免責条項の適用以前に告知義務違反として契約解除や保険金支払拒絶という結果を引き起こすリスクを孕む⁵⁶。シャドーAI と告知義務の関係についてはより立ち入った検討が必要である。告知義務者の知らない事実は原則として告知義務の対象とならない⁵⁷。重過失による不知については告知義務違反が成立しうるが、その重過失はほとんど故意に近似する著しい不注意に限定して解釈すべきとされる⁵⁸。この解釈論を適用すると、経営陣が「現場の従業員が勝手に AI を使っていたとは知らなかった」と抗弁した場合、保険者が契約を解除しあるいは免責を主張することは容易でない。ここで第4節で検討する内部統制システム構築義務が意義を有すると解される。適切な AI ガバナンス体制、すなわち AI インベントリの整備、利用申請・承認プロセスの構築、定期的な監査等が構築されていれば、シャドーAI は容易に検知できたはずである。このような状況では、経営陣の不知は重過失として構成される余地が生じる。内部統制システム構築義務の履行状況が、告知義務違反における重過失認定の基礎となるものとする。

2 AI 除外条項の解釈

第IV章で検討した AI 除外条項が日本の保険契約に導入された場合、その解釈は如何に解すべきか。

AI 除外条項の解釈を検討するにあたり、除外条項の性質決定が前提問題となる。この性質決定論は英国保険法において精緻に議論されてきた問題であり、かつ前章で検討した AI 専用保険の多くがロイズ・オブ・ロンドン市場で引き受けられ英

⁵⁵ シャドーAI とは、企業内で正式な承認を経ずに従業員が個人的に使用する AI ツールをいう。

⁵⁶ Bennett=Corson・前掲注 20) 参照。

⁵⁷ 山下・前掲注 53) 413 頁。

⁵⁸ 山下・前掲注 53) 425 頁。

国法を準拠法とすることに鑑みれば、英国法理を比較法的視点から参照することが有益である。英国保険法理において、保険契約における除外条項は二つの機能を有しうるものとされる。第一に、保険者が引き受けた危険の範囲を画定する機能、第二に、本来カバーされるべき危険を事後的に除外する機能である⁵⁹。前者の場合、条項は保険契約の構成的要素として機能する。後者の場合、条項は保険者の抗弁として機能する。この区別は、契約解釈の方法論にも影響を及ぼす。

第IV章で検討したAI除外条項についてこの枠組みを適用すれば、以下の整理が可能である。AIリスク全般を包括的に排除するAI除外条項は、引受危険の外延を画定する条項としての性格が強い。この場合、次款で検討する約款作成者不利の原則の適用はさらに限定され、文言の平明な意味と契約全体の文脈に基づく通常の契約解釈原則が優先する。他方、特定のAI利用態様のみを免責とする限定的なAI除外条項は、事後的除外としての性格を有することになる。そのため、約款作成者不利の原則を含む約款解釈法理が広く機能する余地がある。AI除外条項の設計の態様が、解釈手法に関係してくることになる。

(一) 約款作成者不利の原則とその限界

日本の裁判実務では、免責条項は保険者による一方的作成であることから、約款作成者不利の原則により制限的に解釈される傾向がある⁶⁰。

もっとも、約款作成者不利の原則の適用には限界がある。山本哲生が指摘するように、解釈によっていずれかの解釈に収斂するのであれば、約款作成者不利の原

⁵⁹ 英国保険法における除外条項の二機能論については、Impact Funding Solutions Ltd v Barrington Support Services Ltd [2016] UKSC 57, paras 6-7 per Lord Hodge 参照。同判決は、除外条項が法律の作用により生じる責任を排除するのではなく保険カバーの範囲を画定する機能を有する場合には、免責条項の制限解釈の一般原則は適用されないと判示した。同判決の法理は、Burnett or Grant v International Insurance Company of Hanover Ltd [2021] UKSC 12において、contra proferentem原則は真の曖昧性 (genuine ambiguity or real doubt) が存在する場合にのみ適用されると整理されている。

⁶⁰ 作成者不利の原則とは、契約条項の解釈に疑義がある場合、当該条項を作成した者に不利に解釈すべきとする原則をいう。不明確条項の解釈方法については、上田誠一郎『契約解釈の限界と不明確条項解釈準則』82-83頁（日本評論社，2003）も参照。

則の適用はない⁶¹。同原則は複数の解釈が成り立ち得る曖昧性が存在する場合にのみ補助的に機能し、文言が一義的に明確である場合には適用されない⁶²。

(二) 合理的期待の原則

作成者不利の原則に加えて、被保険者の合理的期待の保護も重要である⁶³。合理的期待の原則は米国法独自の法理であり、日本法では明文の原則として確立していない。しかし、その実質は日本法においても、作成者不利の原則、信義則（民法1条2項）、および定型約款の不当条項規制（民法548条の2第2項）を通じて考慮されうるものと解される。「～に起因する」といった包括的文言が用いられた場合、AIが単に業務の一環で使用されていたに過ぎない事故まで免責されるとすれば、被保険者の合理的期待を害する。一般の契約者が、専門職賠償保険においてマーケティング資料作成にAIを使用していたことを理由に免責されるとは想定しがたい。日本法の下ではAI除外条項の適用範囲は、損害とAIの挙動との間に相当因果関係が認められる範囲、あるいはAI特有のリスクが直接的な原因となった場合等に限定して解釈すべきと解される。

⁶¹ 山本哲生「作成者不利の原則について」損害保険研究81巻4号（2020年）1頁、4頁。不明確条項の解釈方法については、上田誠一郎『契約解釈の限界と不明確条項解釈準則』82-83頁（日本評論社、2003年）も参照。なお、約款作成者不利益の原則の適用場面が限定的であることを示す判例として、最二小判平成19年6月11日裁判集民224号521頁（セブンイレブン・ロスチャージ事件）がある。

⁶² 英国判例法理でも契約解釈の順序が確立している（英国判例法理における契約解釈アプローチについては、Rainy Sky SA v Kookmin Bank [2011] UKSC 50、Arnold v Britton [2015] UKSC 36、Wood v Capita Insurance Services Ltd [2017] UKSC 24 参照。）。これらの判例は、文言の自然かつ通常の意味を出発点としつつ、契約全体の文脈、商業的常識、当事者の合理的期待を総合的に考慮するアプローチを確立したものと評価される。文言の平明な意味、契約全体の文脈、条項の趣旨、最後の手段としての不利益解釈という段階的アプローチである。除外条項が保険者の引き受けた危険の外延を画定する機能を有する場合には、作成者不利の原則の適用はさらに限定され、通常契約解釈原則が優先する。（Persimmon Homes Ltd v Ove Arup & Partners Ltd [2017] EWCA Civ 373。同判決は、対等な交渉力を有する当事者間の商事契約において、作成者不利解釈は文言の通常の意味および商業的文脈による解釈が結論をもたらさない場合の「最後の手段（last resort）」にすぎないと判示した。）。

⁶³ 合理的期待の原則（doctrine of reasonable expectations）については、Erik S. Knutson, Confusion About Causation in Insurance: Solutions for Catastrophic Losses, 61(5) ALA. L. REV. 957, 967 (2010) 参照。なお、合理的期待の内容は法環境・技術環境の変化に応じて変化するものとする。AI技術が企業活動に遍在化した段階においては、AIを業務に利用することが通常の業務実態となり、AI除外条項への異議申立の根拠として被保険者の合理的期待はより強固なものとなる。

(三) 約款解釈の限界と契約交渉の重要性

もともと、約款の解釈で救済を得ることには限界がある。AI 除外条項の文言が一義的に明確である場合、作成者不利の原則は適用されず、被保険者は免責条項の文言どおりの効果を甘受せざるを得ない。既存契約の約款を事後的に変更することも法的に困難である。保険業法施行規則上の基礎率変更権は、予定発生率に対する実績発生率の状況を示す指標（支払指数）に基づいた行使基準を約款上明確に定めること、毎年保険契約者に対し支払指数の推移を記載した書面を交付すること等の厳格な要件が課されている。保険者にとって事務・システム負荷が大きいため、平成 19 年 4 月以後締結する新契約については基礎率変更権規定を約款から削除する保険会社が多く、同規定は事実上機能していない。また、定型約款の変更（民法 548 条の 4）による対応も考えられるが、変更が相手方の一般の利益に適合する場合、または変更が契約の目的に反せず合理的である場合という要件のハードルは極めて高い。AI 技術の急速な進展に伴いリスクプロファイルが変化したとしても、既存契約に対して AI 除外条項を事後的に挿入することは、被保険者の保険金請求権を制限する不利益変更であり、上記要件を充足することは実質的に困難である。

以上の法的制約を踏まえつつ、米国で最も有効な対抗策として用いられているのは、更改時点で免責の削除や限定を交渉することである。削除が困難な場合には、AI の定義、文言（「に起因する」を「を直接の原因とする」に限定する等）、カーブバック条項⁶⁴の導入を検討する必要がある。消費者契約法 3 条 1 項の透明性原則の観点からは、AI 除外条項の内容が契約締結時に明確かつ平易に説明されることが求められる⁶⁵。

AI 除外条項のような技術的概念を含む条項については、約款の文言のみならず契約概要や注意喚起情報等の募集文書における記載も重要となる。募集文書に明確な説明がなされていれば、被保険者の不意打ちを回避し、合理的期待の内容を形成する上でも意義がある。約款の解釈にあたっては募集文書等の記載も含めて判断すべきであり、募集文書における情報提供の充実は事後的な紛争予防の観点からも重要である。

⁶⁴ カーブバック条項とは、広範な免責条項の中に例外を設け、一定の種類の損害については補償を復活させる条項をいう（例外的復活）。

⁶⁵ 消費者契約法 3 条 1 項は、事業者に対し、消費者契約の条項を定めるにあたっては、その内容が消費者にとって明確かつ平易なものになるよう配慮することを求めている。AI 除外条項のような技術的概念を含む条項については、この透明性原則の要請が特に重要となると考える。

AI 除外条項については、定型約款の組入規制との関係も問題となる。民法 548 条の 2 第 2 項は、定型約款の条項のうち相手方の権利を制限しまたは義務を加重するものであって信義則に反して相手方の利益を一方的に害するものは合意をしなかったものとみなすと定める。AI 除外条項は、被保険者の保険金請求権を制限する条項であるから、同項の適用対象となり得る⁶⁶。もっとも、保険約款は金融庁の認可を経ている以上、実務上は定型約款の不意打ち条項規制による処理よりも、約款解釈による解決がなされる可能性が高いとの指摘がある⁶⁷。この指摘を踏まえれば、AI 除外条項の有効性を確保する上で決定的に重要となるのは、約款の文言自体の明確性に加え、上述した募集文書等における情報提供の充実である⁶⁸。AI 除外条項のように高度に技術的な概念を含む条項が契約者にとって不意打ちとならないためには、契約概要や注意喚起情報において、除外される AI リスクの具体的範囲と残存する補償範囲を平易な表現で明示することが重要である。

3 取締役の義務と AI についての比較法的視座

第IV章で検討したように、海外の AI 専用保険は引受条件として AI ガバナンス体制の構築を求めている。この実務を日本法において法的に基礎づけるには、取締役の AI に関する義務の内容を明らかにする必要がある。AI 時代における取締役会の意思決定と司法審査の関係について、比較法的視座から検討する。

Langenbucher の Ownership and Trust フレームワークは、会社法が取締役会に対して意思決定の所有を期待する一方、経営判断については司法審査を控えるという信頼を付与していることを指摘する⁶⁹。同論文はブラックボックス AI を取締役

⁶⁶ 定型約款の不意打ち条項規制と保険約款の関係については、日本生命保険生命保険研究会『生命保険の法務と実務（第4版）』74頁（きんざい、2023年）参照。なお、不意打ち条項の判断においては、実体的合理性（条項の内容の相当性）と手続的合理性（条項の開示・説明の適切性）の双方が考慮される。

⁶⁷ 日本生命保険生命保険研究会・前掲注66）74頁。もっとも、定型約款制度の創設以降の判例の蓄積は乏しく、今後の裁判例の展開を注視する必要がある。

⁶⁸ なお、私見としては、約款の解釈（平均的当事者概念等）には、約款のみならずご契約のしおり等の記載も取り込んだ上で解釈すべきであると考えられる立場である。理由としては、保険契約は契約であること、ご契約のしおりの定義は約款をわかりやすくしたものであること、ご契約のしおりは約款と一連一体のもの（冊子名称も「ご契約のしおり-定款・約款」ことが多い）として提供される点が挙げられる。また、約款解釈にあたりご契約のしおりに言及する裁判例もある（高知地判平成23年3月23日生命保険判例集23巻126頁）。

⁶⁹ Katja Langenbucher, Ownership and Trust – A corporate law framework for board decision-making in the age of AI, ECGI Law Working Paper N° 758/2025, April 2025.

会の意思決定に使用してはならないという見解を明確に否定している⁷⁰。Langenbucherによれば、AIの予測は既知の未知であり、取締役会が日常的に対処している不確実性の一つに過ぎない。

比較法的には、外部専門家への依拠に関するルールが参考となる。米国デラウェア州会社法 141 条(e)は、社外専門家からの情報については当該専門家が慎重に選任されたことを要求する。ドイツ法においても判例上、外部専門家の意見に依拠する場合、取締役はもっともらしさの確認を行う義務を負うとされる⁷¹。AIをこの枠組みに位置づけるならば、AIベンダーの選定における合理性と、AIの出力に対する批判的な検証が、取締役の善管注意義務の内容となりうる。このフレームワークは、Langenbucherが直接論じてはいないが、AI保険の設計にも示唆を与えると解される。第IV章で検討したMunich Reの技術デューデリジェンスやArmilla AI/Chaucerの引受基準は、取締役会がAIベンダーを慎重に選任しAIの出力のもっともらしさを確認するプロセスを保険という形で制度化したものと理解する。

日本法においても、AI時代における取締役の善管注意義務の内容について検討が始まっている。中村直人は、生成AIが経営判断手続に与える影響を包括的に検討し、経営判断の原則の各要件とAI利用の関係を整理した⁷²。同論考によれば、経営判断の原則の下では、①前提事実の認識に重要かつ不注意な誤りがないこと、②意思決定の過程・内容が著しく不合理でないことが要件とされるところ、生成AIが出力した誤った情報が意思決定の前提事実となった場合、取締役の善管注意義務違反が問題となるものとする。生成AIは現時点では正確性に問題があり、特に非英語圏では誤りが多いとされる中で、AI生成データが意思決定の前提事実に入混入するリスクは無視できない。この問題は、Langenbucherが論じた「AIの予測は既知の未知である」という整理と方向性を同じくする。AIの利用自体は善管注意義務に違反しないが、AI生成情報の正確性リスクを認識した上で適切な手続的統制を講じることが義務の内容となる。

中村論考が特に注目するのは、信頼の原則とAI生成情報の関係である⁷³。判例法理上、大規模な会社においては、意思決定の内部統制がそれなりに整っていることを前提に、取締役は部下からの報告を信頼して自らの意思決定の基礎とすることが許容される。しかし、部下からの報告の中にAI生成情報が含まれている場

⁷⁰ Langenbucher・前掲注 69) 11 頁以下。

⁷¹ BGH, Urt. v. 20. September 2011 - II ZR 234/09, ZIP 2011, 2097 Rn. 18 (「Ision」)。

⁷² 中村直人「生成AIの普及と経営判断手続の見直し」NBL 1245号(2023年)58頁以下。

⁷³ 中村・前掲注 72) 63-64 頁参照。

合、AI の利用について社内で何らルールを定めていなければ、躊躇すべき特段の事情が意思決定のための資料の外観から判別できず、AI 生成情報に含まれる誤りのリスクが見過ごされるおそれがある。この分析は、次節で検討する内部統制システム構築義務としての AI ガバナンス体制構築義務と直結する。AI インベントリの整備、利用ルールの策定、AI 生成情報の明示といったガバナンス体制が構築されていることが、信頼の原則の適用の基盤となる。逆に、ガバナンス体制の不備は信頼の原則の適用を妨げ、取締役の善管注意義務違反のリスクを高めることになる。

さらに中村論考は、AI の信頼性が高まった時代には、AI を意思決定に参照しないことが善管注意義務違反を構成するという可能性を示唆する。情報収集義務の観点から、高度な信頼性を有する AI が利用可能であるにもかかわらずこれを参照しなかった場合には、取締役の情報収集が不合理と評価される余地がある。この指摘は、AI 利用の過少と過剰の双方向にリスクが存在することを意味し、前述の Langenbucher が論じた意思決定の「所有」の維持という要請と、表裏の関係にある。

Kourabas & Tsang の論文はオーストラリア法を素材として取締役会の監視機能と AI の関係を分析している⁷⁴。同論文は、人間が自動化されたシステムに過度に依存しがちである自動化バイアスの問題を指摘し、Human-in-the-Loop アプローチを提唱する⁷⁵。同論文が提唱する具体的施策は、適切な人材を AI 監視業務に配置することと、AI を利用した意思決定の適切な手続を確保することの二点である。

以上の比較法的知見を総合するならば、AI 時代における取締役の義務の核心は、AI を用いた意思決定の「所有」を維持することにある。具体的には、外部の AI ベンダーやシステムへの依拠に際してその選定の合理性と出力の批判的検証を怠らないこと、運用過程において適切な人間の関与を制度的に確保すること、そして

⁷⁴ Steve Kourabas & Cheng-Yun Tsang, The Board Monitoring Function: Artificial Intelligence in the Era of Heightened Accountability, ECGI Law Working Paper N° 856/2025, July 2025.

⁷⁵ Kourabas & Tsang・前掲注 74) 4 頁、29 頁以下。なお、オーストラリア会社取締役協会 (AICD) が 2025 年 12 月に公表した実務報告書は、取締役会自身の AI 利用に関する実務上の「赤信号」(red flags) を整理し、AI 要約への過度の依拠が human in the loop の欠如として善管注意義務違反を構成しうることを指摘する。同報告書が引用する The Star Entertainment Group 訴訟において、Michael Lee 裁判官は「取締役会資料が膨大であることは義務不履行の言い訳にならない」と述べており、AI という読む道具の存在がむしろ注意義務の水準を引き上げる可能性を示唆する点は、本文で述べた中村論考の問題意識と符合する。See Australian Institute of Company Directors, AI Use by Directors and Boards: Early Insights (Dec. 2025).

信頼の原則の適用基盤たる AI ガバナンス体制を構築・維持することがその内実をなす。これらの義務の履行を担保する仕組みとして、AI 保険が機能し得る。

4 内部統制システム構築義務と AI ガバナンス

前節で検討した比較法的知見および第 V 章で検討した課題を踏まえれば、AI 保険が実効的に機能するには企業における AI ガバナンス体制の構築が前提となる。会社法上の内部統制システム構築義務との関係を整理する。

内部統制システム構築義務とは、取締役が会社の業務全部を自ら監視することが事実上不可能であることを前提に、業務の適正を確保するための仕組み・体制を構築する義務である。法的根拠は会社法 362 条 4 項 6 号・5 項等に求められ、大和銀行事件判決以来の判例法理として確立している⁷⁶。

AI 活用に伴うリスクは会社法施行規則 100 条 1 項の各号と対応する⁷⁷。AI システムに起因するシステム停止、不正確な出力、情報漏洩等といった事業上のリスクは損失の危険の管理に関する規程その他の体制（2 号）の対象となる。AI の利用が差別的結果を生じさせる場合や個人情報保護法等の法令に抵触する場合への対応は、使用人の職務の執行が法令及び定款に適合することを確保するための体制（4 号）に関わる。AI の判断過程・学習データ等の記録・保存は、取締役等の職務の執行に係る情報の保存及び管理に関する体制（1 号）の問題となる。子会社等が AI を利用する場合のグループ全体でのガバナンス方針の統一は、企業集団における業務の適正を確保するための体制（5 号）に関わる。

以上から、取締役には内部統制システム構築義務の一内容として AI ガバナンス体制構築義務が認められると解される。日本銀行金融研究所の報告書も、取締役は内部統制システム構築義務の一内容として AI ガバナンス体制構築義務が認められ、AI の利用について組織内の最終的な責任を負うと解されると述べている⁷⁸。

同報告書は、AI ガバナンス体制構築義務の具体的内容として、組織面の整備、リスクベースのモニタリング、説明責任の確保、および定期的なアップデートという枠組みを提示する⁷⁹。組織面では、AI のリスクに関する社内教育の徹底・社内

⁷⁶ 大阪地判平成 12 年 9 月 20 日判時 1721 号 3 頁（大和銀行事件）。

⁷⁷ 会社法施行規則 100 条 1 項参照。

⁷⁸ 金融機関における AI の利用を巡る法律問題研究会『金融機関における AI 利用に伴う私法上のリスクと管理』（日本銀行金融研究所、2025 年 6 月）。

⁷⁹ 金融機関における AI の利用を巡る法律問題研究会・前掲注 78) 参照。同報告書は金融機関を主たる対象とするが、AI ガバナンスの組織的基盤に関する枠組みは一般事業会社にも妥当し得る。なお、金融庁「モデル・リスク管理に関する原則」（2021 年 11 月）も、AI の利用方

ルールの整備、専門部署によるリスクの定性的・定量的評価、リスクレベルに応じた経営陣の関与体制の整備が求められる。かかる組織的基盤を前提として、リスクベースのモニタリングとして、想定されるリスクの特定と評価、リスクの大きさに応じた頻度のモニタリング、リスクが顕在化した場合の異議申立権の付与や事後的な問題説明機会の提供等が必要となる。これにとどまらず、AI で利用しているデータやモデルの透明性・公平性を確保し公表するという説明責任の履行も不可欠である。技術の急速な進展に照らせば、データやモデルの更新の要否を定期的に確認し外部環境の変化を適時に反映する不断の見直しもまた、義務の内容に含まれると解される。

ここで注目すべきは、内部統制システム構築義務の水準に関する同報告書の指摘である。同報告書は、不適切な行為が行われた時点の水準に照らし、通常想定されるリスクにつき、同業他社並みの水準で体制を構築する必要があるとした上で、AI 技術は進展が速いため AI ガバナンス体制構築後も不断の見直しを行うこと、グローバルに展開する企業においては海外同業他社の水準を意識することが求められると述べる⁸⁰。この指摘は、AI ガバナンス体制の善管注意義務上の水準が静的なものではなく、技術進化に応じて動的に変化し得ることを示唆しており、前節（第 3 節）で検討した Langenbucher や Kourabas & Tsang の比較法的分析とも関連する問題意識である。もっとも、同報告書の枠組みは主として AI ガバナンスの組織的基盤に関わるものである。生成 AI が従来の AI システムとは異なる固有の特性を有することに鑑みれば、技術的統制の具体的内容についてさらなる検討が必要となる。

法や態様に応じたリスク評価と AI 特有のリスクの考慮の必要性を指摘する。AI ガバナンス体制の具体的内容については、英国取締役協会（Institute of Directors）の調査を踏まえて Burges Salmon が 2025 年 11 月に公表した報告書が 12 原則を提示しており、取締役会レベルでの AI 統治の実務指針として参考になる。See Charlotte Hamilton & Eleanor Furlong, AI Governance in the Boardroom: Guidance for Business Leaders, Burges Salmon (Nov. 17, 2025).

⁸⁰ 金融機関における AI の利用を巡る法律問題研究会・前掲注 78) 参照。内部統制システム構築義務の水準については、大和銀行事件判決（前掲注 76）参照）が、当該行為がなされた当時の状況および当該会社の属する業界における通常の経営者の有すべき知見・経験を基準として判断すべきことを示している。AI 技術の急速な進展は、この基準自体が継続的に変動することを意味する。

EIOPA が 2025 年に公表した生成 AI 市場調査の知見がここで重要な示唆を与える⁸¹。同調査は、生成 AI のガバナンスにおいて推論段階への統制シフトが生じていることを指摘する。従来の機械学習では、企業は自社でデータを収集しモデルを学習させ、その開発過程を内部で検証・監査できた。ガバナンスの焦点は学習データの品質管理、モデル開発プロセスの統制、定期的な再学習の実施等にあった。これに対し生成 AI では、基盤モデルがビッグテック等の第三者によって事前学習されており、企業は当該モデルを API 経由で利用するか RAG・ファインチューニング等を施すにとどまる場合が多い。モデル学習段階への直接的な統制は限定的であり、AI ガバナンスの主戦場は推論段階へ移行する。以上の組織的な基盤を前提として、生成 AI の推論段階における技術的統制の内容は、以下のように整理できる⁸²。

入力統制の観点からは、生成 AI システムへのプロンプト入力で何を許容するかを明確にし、プロンプト・インジェクション攻撃や機密情報の意図しない入力を防止する体制の構築が求められる。これと表裏の関係にあるのが出力監視であり、特にハルシネーションの検出と対処の仕組みの整備が必要となる。顧客対応に AI を用いる場合には、出力が募集規制上の情報提供義務等の法令に適合するかの確認体制も不可欠である。Agentic AI の導入に際しては、こうした入出力の統制に加え、外部システムへのアクセス権限の範囲を明確に画定し AI が自律的に行使し得る権限の上限を設定する権限設計が求められる。さらに、以上の統制を実効的に担保するのがログ管理であり、AI の入出力履歴および人間による介入・承認の記録を保存し説明可能性と責任の追跡を担保する体制は、保険事故発生時の因果関係立証や規制当局による監督対応においても決定的な意義を有する。これらの統制は会社法施行規則 100 条 1 項各号の内部統制体制と対応し、入力統制・出力監視は 4 号に、ログ管理は 1 号に、権限設計は 2 号にそれぞれ関わる。

⁸¹ EIOPA・前掲注 8) 27 頁。同報告書は、「生成 AI システムについては、モデル学習段階よりも推論段階（プロンプト・エンジニアリングおよび出力監視等）により重点を置く必要がある」と指摘する。

⁸² この統制枠組みは、EIOPA 調査の知見を踏まえて筆者が整理したものである。同調査は、生成 AI のリスク管理において、従来の AI システムで用いられてきたモデル学習段階の検証技術の多くは、生成 AI には適用できないとし、推論段階での統制強化の必要性を指摘する（EIOPA・前掲注 8) 27-28 頁）。なお、入力統制・出力監視・記録保存の各項目は同調査 30-31 頁に依拠する。その余は筆者の考察による。

5 保険による規律付けの可能性

前節で検討した AI ガバナンス体制構築義務と保険はいかなる関係に立つか⁸³。

AI 分野では法規制、ハードローの整備が技術進化に追いつかない傾向が強い。第 I 章で触れた王 (2024) が整理する「保険＝ガバナンス」理論によれば、保険者は引受決定や契約上の予防措置の要求を通じて被保険者の行動に影響を与え、社会的リスクガバナンスにおいて第三者による監督の機能を果たす⁸⁴。この理論的視座に照らせば、保険会社が AI 専用保険の引受条件として厳格なガバナンス体制の遵守を求めることは、事実上のソフトローとして機能する。小塚は、保険商品の前提条件として AI 開発者等に AI 原則の実施や AI ガバナンスの構築を義務づけることで、「AI リスク保険が普及するとともに AI ガバナンスも関係業界に浸透するという好循環が期待できる」と論じている⁸⁵。

第 IV 章で検討した Munich Re のパフォーマンス保証保険や Armilla AI/Chaucer の Affirmative AI Liability Insurance は、小塚が論じた契約関係のガバナンスを保険で補完・強化するものと位置づけられる⁸⁷。Chaucer は故障に陥りやすい AI の引受けを選別する方針を明示しており、保険会社が AI ガバナンスのゲートキーパーとして機能する実例が既に存在する。今後は、保険会社がデューデリジェンスを行う際の基準として、NIST AI RMF や ISO/IEC 42001 への準拠を保険引受の条件とする動きが出てくる可能性がある⁸⁸。米国 NAIC が 2023 年に採用したモデル通

⁸³ 前掲注 6) 参照。

⁸⁴ 王・前掲注 4) (中)13 頁、20 頁。王論文は、Rappaport (2017) が現代の保険制度において伝統的な事後損失補償メカニズムから事前損害防止・抑止メカニズムへの移行が生じていると論じたことを紹介する (同 13 頁)。さらに、保険者が法定義務を保険契約条項の解釈作業に組み込むことで被保険者の義務履行を事実上監督する「監督者」となると指摘する (同 20 頁)。

⁸⁵ 小塚・前掲注 6) 32 頁。

⁸⁶ Trout の論文は、保険が規律付けとして機能し得る条件を理論的に整理し、フロンティア AI 分野への適用可能性を論じている。(Cristian Trout 「When Does Regulation by Insurance Work? The Case of Frontier AI」 arXiv (2025 年 12 月 6 日公表))

⁸⁷ 小塚・前掲注 6) 参照。

⁸⁸ NIST AI RMF は、米国国立標準技術研究所が公表した AI リスク管理のためのフレームワークであり、ISO/IEC 42001 は、AI マネジメントシステムに関する国際規格である。

達は、保険会社自身による AI システムのガバナンス・文書化・監査手続を求めるものであり、日本においても参照される可能性がある⁸⁹。

これにより、法規制というハードローから標準化というソフトロー、そして保険という市場メカニズムへと連なるエコシステムが将来的に形成される可能性がある。技術デューデリジェンスを通じて企業の AI ガバナンスを点検し、リスクの低い企業に料率優遇を行うことで、産業界全体の安全性を底上げすることが期待される。これは保険業法 1 条が予定する保険の公共性とも整合する⁹⁰。ただし、免責条項の設計が不十分な場合には逆効果にもなりうる。保険がガバナンスを促す規律として働くには、免責と補償の境界が企業の AI ガバナンス体制を統制する企業努力と整合的である必要がある。

EIOPA 調査によれば、EU 保険市場において専用 AI ポリシーを策定している保険会社は 49%に達し、2023 年の 25%から 2 倍に増加している⁹¹。保険会社自身が AI ガバナンスの重要性を認識し自らの業務運営において実践していることを示している。AI 保険の引受者たる保険会社がまず自社において AI ガバナンスを実践することは、被保険者にガバナンス体制の構築を求める際の説得力を高め、保険による規律付けの実効性を担保する上でも重要である。

もともと、保険による規律付けには実務上の限界がある。保険法 28 条は、保険媒介者による告知妨害および不告知教唆があった場合、保険者は告知義務違反を理由とする契約解除ができないと定める。代理店・ブローカー等の媒介者が「これは書かなくていい」等と勧めた場合、保険者の解除権は阻却される⁹²。AI 保険の販売において、この規定は重大なリスクとなる。技術的理解の浅い代理店やコンサルタントを兼ねるブローカーが、「この程度の AI 利用なら告知書に書かなくても大丈夫」「シャドーAI まで申告する必要はない」といった安易な助言を行えば、それは不告知教唆に該当する可能性がある。こうした場合、保険者は告知義

⁸⁹ NAIC, Model Bulletin: Use of Artificial Intelligence Systems by Insurers (Dec. 4, 2023). 同通達は、保険会社に対し AI システムの責任ある利用のための書面プログラム (AIS Program) の策定・実施・維持を求め (Section 3)、規制当局による検査時に当該プログラム等の情報・文書の提出を求めうる旨を定める (Section 4)。

⁹⁰ 保険業法 1 条は、「保険業の公共性にかんがみ、保険業を行う者の業務の健全かつ適切な運営及び保険募集の公正を確保することにより、保険契約者等の保護を図」ることを目的として掲げている。

⁹¹ EIOPA・前掲注 8) 25-26 頁 (Figure 13)。同調査によれば、生成 AI を利用している保険会社の 69%が既に AI ポリシーを策定しており、さらに 16%が 3 年以内に策定予定である。

⁹² 山下・前掲注 53) 436 頁以下。

務違反を理由とする解除ができず、規律付け機能は機能しえない。AI 保険による規律付けの実効性を担保させるには、保険会社自身による直接的なデューデリジェンス、または販売チャネルに対する AI リテラシー教育が必要となる。

6 Agentic AI と法的責任の帰属

今後の展望として、保険と自律型エージェント (Agentic AI) に関する検討が必要となる⁹³。従来の生成 AI が情報の生成や要約を主たる機能としていたのに対し、Agentic AI は自律的に推論し、計画を立て、システムを操作してタスクを完遂する能力を持つ。ここで生じるのはエージェントの行為の法的効果の帰属問題である。第IV章で検討したエア・カナダ判決が示唆的である⁹⁴。民法上の原則に照らせば、AI を業務に利用している以上、その行為の効果は原則として企業に帰属すると解される。履行補助者としての過失や使用者責任の類推が考えうる⁹⁵。しかし、AI の自律性が高まり人間の具体的な指図を超えて行動した場合、その予見可能性の範囲をどう認定するかは難題である。かかる問題は第4節で検討した内部統制システム構築義務と関連する。

VII むすびにかえて

本稿は、サイバー保険の発展過程から得られた知見を手がかりに、AI 保険の法的課題を比較法的に検討した。第II章ではAI リスクの特性を類型化し、従来のサイバー保険の枠組みでは対応しきれないことを示した。第III章では Merck 事件 (戦争免責条項のサイバー空間適用) および G&G Oil 判決 (商業犯罪保険のコンピュータ詐欺条項のランサムウェア損害への適用可能性) の分析を通じて、約款の精緻化と補償範囲の明確化という教訓を抽出した。第IV章では海外における AI 除外と AI 専用保険の二極化を検討した。この動向は AI リスクをサイレント・カバレ

⁹³ Agentic AI (自律型エージェント) とは、自律的に推論し、計画を立て、タスクを完遂する能力を持つ AI をいう。

⁹⁴ *Moffatt v. Air Canada*, 2024 BCCRT 149. 前掲注 31) 参照。

⁹⁵ AI を業務に利用する企業が行為主体として民事責任を負うという帰属構造については、日本銀行金融研究所「金融機関における AI の利用を巡る法律問題研究会」が「AI を法的主体としては扱わず、もっぱら AI を利用している主体の過失の問題として扱う」立場を明示する (日本銀行金融研究所「金融機関における AI 利用に伴う私法上のリスクと管理」(2025年6月) 18頁注 52 (https://www.imes.boj.or.jp) (2026年3月12日最終閲覧))。同報告書は、契約責任・不法行為責任のいずれの局面においても、AI がサービスの提供主体たる企業の過失問題として処理される旨を一貫して論じており (同報告書 16-21頁)、履行補助者の過失や使用者責任の類推による帰属論と整合的である。

ッジから切り出し、AI の失敗様式に合わせて補償を再設計する保険市場の合理的な対応である。第V章ではAI 保険確立に向けた課題として、リスク評価の困難性、因果関係論の限界、パラメトリック保険の可能性を論じた。パラメトリック保険の保険法上の位置づけについては、吉澤卓哉の定額給付型損害保険契約に関する分析枠組みを援用し、被保険利益が存在し、パラメータと損害との間に合理的な相関関係が認められ、賭博禁止とモラル・ハザード防止が確保される限り、損害保険契約としての性質を維持し得ることを論じた。第VI章では日本への含意として、告知義務とAI インベントリ、AI 除外条項の解釈、内部統制システム構築義務としてのAI ガバナンス体制構築義務、保険による規律付けの可能性を検討した。EIOPA 調査に基づき、生成AI のガバナンスにおいて推論段階への統制シフトが生じているという知見を踏まえ、AI ガバナンス体制構築義務の具体的内容として、日本銀行金融研究所報告書の組織的ガバナンス枠組みを基礎としつつ、生成AI 固有の推論段階における技術的統制として入力統制・出力監視・権限設計・ログ管理の四点を提示した。比較法的視座として、Langenbacher の Ownership and Trust フレームワーク、中村直人の経営判断手続と信頼の原則に関する分析、Kourabas & Tsang の Human-in-the-Loop アプローチを検討し、AI 時代における取締役の義務の内容を整理した。

AI 保険の確立には、サイバー保険の教訓を踏まえた約款の精緻化と、内部統制システム構築義務の一内容としてのAI ガバナンス体制構築義務の確立が不可欠である。そして、AI 保険はこのAI ガバナンス体制の構築を促す規律付けの機能を担い得る。

パラメトリック保険の保険法上の位置づけについては、第V章で吉澤の定額給付型損害保険契約論を援用して理論的整理を試みたが、AI 保険に特有のモデル精度、システムダウンタイム等のパラメータ設計と損害との相関関係の実証的検証は今後の課題である。AI 除外条項の解釈に関する裁判例の蓄積を待って、より精緻な分析を行う必要がある。保険による規律付けの実効性について、実証的な検証も求められる。生成AI の社会実装が進む中、ブラックボックス化する因果関係の法的構成、約款におけるAI 定義の精緻化、保険の社会的規律付け機能について、さらなる研究が求められている。翻って本稿の分析が示すのは、AI 保険の確立は単なる新商品の開発にとどまらず、保険がAI ガバナンスの規律付けの仕組みとして機能し得るという点である。サイバー保険が「保険＝ガバナンス」理論のもとで情報セキュリティ管理の産業界への浸透を促す機能を担ってきたと論じら

れているように⁹⁶、AI 保険も AI ガバナンスの普及を促す制度的な基盤となると考
える。その実現に向けた法的条件の整備が課題であるといえる。

本稿が今後、産学の連携をより一層深めた AI 保険の法的基盤の形成に向けた議論
の出発点となることを期待しつつ本稿を閉じることとする。

⁹⁶ 王・前掲注 4) (中)17-18 頁。同論文は、サイバー保険が被保険企業の「コンプライアンス
マネージャー」として機能し、引受条件においてサイバーセキュリティ基準の遵守を求めるこ
とで、関連法規の実行を事前に確保する機能を担ってきたことを論じる。