IFI Policy Recommendation

IFI Policy Recommendation No.24 October 2023

AI ガバナンスに資する AI 監査の実践に向けて

東京大学未来ビジョン研究センター
技術ガバナンス研究ユニット
AI ガバナンスプロジェクト AI 監査研究会

東京大学未来ビジョン研究センター
Institute for Future Initiatives
The University of Tokyo

要旨

人工知能（Artificial intelligence: AI）が社会の様々なサービスやシステムに組み込まれて利用されるにあたり、多くの企業や組織が AI 原則やポリシー、あるいはコミットメントを提案している。一方、AIサービスやシステムを開発、提供側による自主的な原則だけではリスク対応には不十分であるとし、独立した監査の必要性も提唱されている。本稿では、AIサービスやシステムの監査に関する論点を整理して、AI ガバナンスに資する AI 監査を目指すための展望を示す。

AI監査をめぐる論点は多岐にわたり、また同じ単語であっても立場や前提の違いによって想定する内容が異なるため議論のすれ違いが発生しやすい。そのため、AI監査の実践に向けては、監査対象や監査のタイミングなどの論点についてそれぞれ共通理解に基づいて関係者が議論をすることが重要である。

本稿の第 2 章においては AI 監査を議論する上で想定される(1)監査の必要性、(2)立証命題、(3)監査対象、(4)実施タイミング、(5)実施者要件、(6)監査関係者と関係組織の 6 つの論点を取り上げ、それぞれについて概観を整理した。また第 3 章では、何故 AI 監査の実施が困難であるのかについて技術的、制度的、社会的等の様々な観点から検討した。これらの監査上の論点、監査を困難とする要因について、具体的なイメージが持てるよう第 4 章では採用 AI 事例を想定した検討も併せて実施した。

以上をふまえ、第 5 章では AI 監査を今後適切に進めて行くために以下の 3 点を提言した。

提言 1. AI 監査の制度設計の整備
提言 2. AI 監査に関する人材の育成
提言 3. 技術や利用の進展に伴う AI 監査のアップデート

なお、本稿では入力データに対する認識や予測を行う AI を想定して検討を進めているが、生成 AI をめぐる論点についても稿末の補章にて紹介している。

AI 監査研究会について

本政策提言は東京大学未来ビジョン研究センター技術ガバナンス研究ユニットの一プロジェクトである AI ガバナンスプロジェクトの研究会である「AI 監査研究会」の成果報告である。AI 監査研究会は 2022 年 6 月より研究活動を開始し、以下のメンバーより構成される。

江間有沙（東京大学未来ビジョン研究センター・准教授）
佐藤　亮（東京大学未来ビジョン研究センター・客員研究員/有限責任監査法人トーマツ）
長谷友春（東京大学未来ビジョン研究センター・客員研究員/有限責任監査法人トーマツ）
中野雅史（東京大学未来ビジョン研究センター・客員研究員/東洋大学）
上村信二（有限責任監査法人トーマツ）
北村　弘（CDLE・AI リーガル（日本電気）/IRCA（International Register of Certificated Auditors：国際審査員登録機構）「ジャパン」メンバーズサポーター）

本政策提言策定のプロセス

本政策提言の内容は東京大学未来ビジョン研究センター技術ガバナンス研究ユニットの AI ガバナンスプロジェクトメンバーと外部有識者からなる AI 監査研究会の議論を取り纏めたものである。研究会は 2022 年 6 月より、月 1 回程度の頻度でオンラインミーティングを重ね、有識者の方々へのヒアリングも行って、本提言を作成した。本稿を取りまとめるにあたってヒアリングにご協力いただいた方々のリストは謝辞を参照されたい。

また、本政策提言は今後、関係者からのコメントや 2023 年 11 月に予定されているウェビナーや他関連イベントからのフィードバックを有効に活用して改訂および充実を図る予定である。

目次

1. AI ガバナンスに資する AI 監査を目指して

人工知能（Artificial intelligence: AI）が社会の様々なサービスやシステムに組み込まれて利用されるにあたり、多くの企業や組織が AI 原則やポリシー、あるいはコミットメントを提案している。一方、AI サービスやシステムを開発、提供側による自主的な原則だけではリスク対応には不十分であるとし、独立した監査の必要性も提唱されている。本稿では、AI サービスやシステムの監査に関する論点を整理して、AI ガバナンスに資する AI 監査を目指すための展望を示す[1]。

1.1 AI 監査の論点整理の必要性

AI の監査対象やタイミング等の議論は多岐にわたるため、共通理解がなければ同じ AI 監査という単語を用いても議論のすれ違いが発生する[2]。例えば監査対象となる AI 技術が機械学習を想定しているのか、それともより狭義に深層学習を想定しているのかという AI 技術の観点で前提の確認が必要である。また、AI 技術以外にも学習データセットや採用されている外部ライブラリ等も監査の対象と考えるのかと言った、対象スコープについても人によって想定が異なる。その他にも、監査を実施するタイミングが開発段階か、それとも利活用段階なのか等、議論の前提となる論点には様々なものが考えられる。なお、本稿においては深層学習を含む機械学習全般を対象として議論を進めていくが、入力データに対する認識や予測を行う AI を想定している。入力データや指示に沿って新しいコンテンツやデータを生み出す生成 AI をめぐる論点については稿末のコラムにて紹介する。

監査には金融証券取引法に基づく財務諸表監査のような法令に基づく法定監査と、法的拘束力のない任意監査に分類でき、任意監査はさらに様々なフレームワークにより実施されている。日本では現在、AI に特化した法定監査は存在しない。しかし他の法令等に基づく監査を実施する際に、AI サービスやシステムが監査対象となる可能性がある。例えば被監査企業が AI システムを用いて財務諸表の数値を作成したり、AI システムの判断結果に依拠した内部統制を構築したりしている場合、その AI システムも監査対象となりうる。具体的には、財務諸表監査や内部統制監査で重要性があれば AI に関連する内部統制の整備運用状況等が検証されることとなる。

---

[1] AI と監査をめぐる論点には（1）AI サービスやシステムなどの監査、(2)AI サービスやシステムを監査手続に使う、(3)監査人と監査業界における仕事の未来の議論がある（中野, 2023）。本稿は(1) AI を対象とした監査を対象としており、(2)監査の手続ツールとして AI サービスやシステムを用いることは直接的には議論の対象とはしない。なお、(2)と(3)の論点に関して先駆的な研究には(Issa et al., 2016)がある。

[2] AI の定義は専門家間でも一意に定まっていないが、本稿では OECD の「AI に関する専門家会合（AIGO）」による以下の AI システムの解説に基づいて論点の整理を行っている。「人間が定義した一定の目的のために、現実または仮想の環境に影響を及ぼす予測、提言または判断を行うことができる機械ベースのシステムである。AI システムは機械もしくは人間またはその双方によるインプットを基に現実もしくは仮想の環境を認識し、そのような認識を（機械学習を用いる自動化された方法または手動で）モデル化し、モデル推論を基に情報や行動の選択肢を定式化する。AI システムは様々なレベルの自律性をもって動作するように設計されている。」（https://doi.org/10.1787/d62f618a-en）

それ以外のAIサービスやシステムを監査する場合は、内部監査や外部監査であっても任意監査となり、監査目的や対象、タイミング等を監査人が個別に設計する必要がある。そのため、なぜ監査をする必要があるのか、何を対象としていつ誰が監査するのかを関係者間で共有することがAIガバナンスに資するAI監査の実現に必要である。

1.2 AI監査に係る二つの潮流

AIサービスやシステムに対する監査には、理論を中心としたアカデミズムでの議論と実務を中心とした実務界での議論の大きく二つの潮流があり、本稿ではこの二つの潮流から論点を整理する。

一つ目は、AI技術の倫理やガバナンスについて議論する学際的なコミュニティにおいて、AIサービスやシステムの監査を、信頼してAIサービスやシステムを開発、利活用するための手段と位置づけた議論である。もう一つは、内部監査や外部監査に係る監査人における実務的な観点からAIサービスやシステムの監査を実施したり、その難しさなどを議論するものである。

本稿ではこの二つの潮流に係る有識者等にヒアリングも行いながら論点を整理した。

1.3 本稿の章立て

本稿では2章でAIサービスやAIシステムの監査の必要性、立証命題、対象、タイミング、実施者要件、関係者と関係組織という6つの側面を整理する。続く3章ではAI監査を困難にしている技術的、制度的、社会的な原因を検討したのち、2章と3章を人事採用時に利活用される採用AIに当てはめた事例を4章で紹介する。5章ではAI監査に対する今後の課題と提言を示し、6章でまとめを行う。最後、7章ではコラムとして2022年以降急速に利用が拡大し始めた生成AIに対する監査で考えるべき論点を紹介する。

2. AI監査をめぐる論点

本章では、AI倫理やAIガバナンスをめぐる論点からAI監査の必要性を整理したのち（2.1）、監査における立証命題(2.2)、対象(2.3)、タイミング(2.4)、実施者要件(2.5)と関係者と関係組織(2.6)の論点を概観する。

2.1 AI倫理やAIガバナンスの観点からみるAI監査の必要性

2010年代後半以降、AIサービスやシステムの利用が拡大していく中、AI倫理やガバナンスに関する議論が産学官民様々な関係者で議論されてきた。

(1) AI倫理の観点から必要とされる監査

AI原則には、透明性、公平性、安全性など様々な価値が議論されている（Jobin et al 2019）。AIサービスやシステムがこれらの原則に則していると保証してもらうため監査が必要であるとする考え方がある。AI倫理に関する議論においては、監査は第三者認証や標

準化など他の手法と並列で議論されることも多く、利用者がAIシステムやサービスを安心して使うための枠組みの一つと捉えられている[3]。また、AI サービスやシステムには様々なリスクが想定されているため[4]、AI システムの提供側も、利用者が監査による保証を求めるのであれば監査を受けるインセンティブが設定される。

国内あるいは国際標準により、組織や国をまたいで開発、運用、利用が行われるAIシステムにおいては倫理原則を実践に落としていくにあたって、リスク管理や監査における枠組み間の相互比較、相互運用が求められる。この枠組み間の「相互運用性」は、2023年のG7 デジタル・技術閣僚宣言の附属書 5 で「AI ガバナンスの枠組み間の相互運用性」との言葉でも紹介されている[5]。国内プロセスの相互調整を伴う相互承認や十分性認定とは異なり、「枠組み」というレイヤーで相互運用性を促進することにより、各国・各組織のAIに対する規律や対応が並存、協調できるようになる[6]。そのためには AI に関連する用語、基本概念やAIガバナンス・マネジメント等に関する各種の標準の議論とコンセンサスが必要

[3] 日本の統合イノベーション戦略推進会議の下で開催されているAI戦略会議の論点整理においても、AI開発者やサービス提供者が現行法令やガイドラインに則り情報開示を求めていること、透明性や信頼性の確保をすること、その際に第三者認証制度や監査制度も参考とすべきと書かれている（https://www8.cao.go.jp/cstp/ai/ronten_honbun.pdf 、p.10）。

[4] 例えば総務省「AIネットワーク化検討会議報告書2016」では（A）AIに期待される機能が適正に発揮されないリスク（セキュリティリスク、透明性・説明可能性のリスク・制御喪失のリスク等）と、（B）AIにより権利利益等法益が侵害されるリスク（プライバシー・個人情報保護に関するリスク、犯罪に使用されるリスク、消費者等の権利利益に関するリスク、人間の尊厳と個人の自律に関するリスク、民主主義と統治機構に関するリスク）に分類されている（https://www.soumu.go.jp/menu_news/s-news/01iicp01_02000050.html）。

[5] G7 群馬高崎デジタル・技術大臣会合の開催結果、附属書5「AIガバナンスのグローバルな相互運用性を促進等するためのアクションプラン」。https://www.digital.go.jp/news/efdaf817-4962-442d-8b5d-9fa1215cb56a/#declaration

[6] 例えば、OECDがISOやIEEE、NISTの標準、欧州AI規制法案、欧州評議会の影響評価法（HUDERIA）等を「相互運用枠組み（interoperability framework）」として項目比較を行っている。OECD.AI work promoting interoperability of AI risk management frameworks, IGF Policy Network on AI meeting #4, 18 July 2023,　https://www.intgovforum.org/en/filedepot_download/282/25999
OECDは他にもAIシステムの分類や、AIシステムのライフサイクルを分類し、比較できる「枠組み」を提供している（https://doi.org/10.1787/2448f04b-en）。日本では経済産業省が「AI原則実践のためのガバナンス・ガイドライン」を公開している。
https://www.meti.go.jp/press/2021/01/20220125001/20220124003.html

であり、現在 ISO/IEC[7]や IEEE[8]、NIST[9]、CEN-CENELEC[10]等の標準化団体によって標準化が推進されている。複数の規律による二重管理はイノベーションを阻害し、結果として社会全体としての損失につながる可能性があるため、これらの枠組みを各国・地域や AI システムの応用分野で適切に運用できる仕組みづくりも求められている[11]。

(2) AI ガバナンスの観点から求められる監査

AI 技術の進展が早いことにより、既存制度やガバナンス(統治)の仕組みでは対応できないことも多い。そのためアジャイル（機動的で柔軟な）に制度やシステムのアップデートを行うアジャイル・ガバナンスが提案されている。機動的なガバナンスシステムの信頼性を評価するためのアプローチの一つとして、内部監査や外部監査などが言及される[12]。

AI サービスやシステムの特徴の一つとして、実装後も再学習により定期的なアップデートを行うことが挙げられる。その際、適切に運用されているかの監視や、不具合が生じたときには保守を行う必要がある。このような適切な管理を受けられない AI サービスは精度が落ち、誤判定が起きるなど社会的なリスクを増幅させる可能性がある。このように、AI サービスやシステムの構築段階のガバナンスの有効性に加えて、運用・保守フェーズにおけるガバナンスの有効性という観点からも監査が必要とされる。

さらに、AI 監査の対象となるサービスごとに、求められる監査の水準と監査にかかるコストのバランスについても検討する必要がある。アプローチの一つとしてリスクベースアプローチが挙げられるが、これは対象の AI サービスに関連するリスクの高低に応じて必要な監査の水準を決定し、その水準に応じて手続及びコストの濃淡をつけて監査を実施するものである。例えば欧州 AI 規制法案はリスクベースアプローチを採用しており、ハイリス

---

[7] ISO/IEC JTC 1/SC42 では、これまでに 人工知能の概念及び用語（ISO/IEC 22989）や AI（人工知能）の利活用が組織のガバナンスに与える影響（ISO/IEC 38507）等、すでに 20 の国際標準が作成・公表されており、また 31 の国際標準化が議論中である。

[8] IEEE の 7000 シリーズ他では AI の実務上の課題を対象とした規格が議論されている。

[9] 米 NIST は ISO/IEC のマネジメント基準や概念、OECD の AI 勧告と相互運用可能な AI の統一リスクベースフレームワークを議論し、関係性を整理している。

[10] CEN/CENELEC では欧州における AI の規格が議論されている。

[11] この枠組み間の相互運用性などに関する議論に関しては、東京大学未来ビジョン研究センターが公開した政策提言「AI の責任ある展開に向けて：広島 AI プロセスへの政策提言」（https://ifi.u-tokyo.ac.jp/news/16642/）を参照。また東京大学未来ビジョン研究センターでは、AI サービスやシステム監査の前段階の作業等としても活用できる、リスクチェーンモデルという AI ガバナンスのフレームワークを提供し、そのケース事例をウェブサイトで公開している。リスクチェーンモデルのウェブサイトとケースは以下で公表されているほか（https://ifi.u-tokyo.ac.jp/projects/ai-service-and-risk-coordination/）、政策提言「AI サービスのリスク低減を検討するリスクチェーンモデルの提案」は以下で公開されている（https://ifi.u-tokyo.ac.jp/news/7036/）。

[12] 経済産業省「アジャイル・ガバナンスの概要と現状」（2022）等でも議論されている。
https://www.meti.go.jp/press/2022/08/20220808001/20220808001.html

ク AI システムに対しては独立した監査人の報告を含める必要があるとしている[13]。

2.2 AI 監査の立証命題

監査を実施する際に何を監査の主題とするかを、本稿では監査の「立証命題」という用語を用いて整理する[14]。AI 監査の立証命題としては、AI 原則に登場する項目や、各種指針やガイドラインで頻繁に登場する項目、AI 以外のシステムの監査でも立証命題として検討する項目等が考えられる[15]。これらの項目のうち、AI サービスやシステムに求められる監査の立証命題のなかで特に AI 監査に特徴的なものとしては、例えば表 1 のようなものが挙げられる。

表 1　AI 監査の立証命題例

立証命題	解説
公平性	AI システムの出力結果に不適切なバイアスがかかっていないか　等 公平の定義についても予め共通認識を持つ必要がある
透明性	AI システムの出力結果について再現ができるか、学習データや採用されている特徴量（パラメータ）について説明が可能か　等
安全性	AI システムが利用者に危害を加える可能性はないか、不具合が発生した場合に適切に停止状態に移行するか　等 AI システムが組み込まれているハードウェアについても考慮する必要がある
セキュリティ	学習データに対する攻撃を予防・発見できるか、意図的に不適切な出力を誘導するような本番入力データを予防できるか　等
プライバシー	個人が共有を望まない属性データを拒否できるか、誤った個人評価を適時適切に訂正できるか　等

[13] Council of the European Union, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - General approach (2021), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

[14] 「監査の対象」という用語を用いると、2.3 に列挙している監査として証明していく個別命題と混乱が生じるため、ここでは監査の「立証命題」という用語を用いる。なお、鳥羽（2009）は監査の対象を監査の「主題」という概念で切り離し、監査プロセスにおいてさまざまな形で独特に登場する監査対象と峻別している。

[15] 総務省「AI ネットワーク社会推進会議 報告書 2022」(https://www.soumu.go.jp/main_content/000826564.pdf)では、各国原則・指針・ガイドラインをもとに 22 の尊重すべき価値を確認している。

これらの立証命題に対して監査を実施する場合、既存のシステム監査で用いられる基準・規準やフレームワークの一部は応用可能である。例えばセキュリティやプライバシーについては既存の SOC2 報告書で採用されているトラストサービス規準を用いて一定程度カバーすることが可能と考えられる[16]。一方、AI の技術的な複雑性や広範な関与者、AI ならではの考慮事項を加味すると（3.1 参照）、既存の基準だけで AI 監査における立証命題を全て対応するのは難しい。

2.3 AI 監査の対象

AI システムやサービスの監査対象は、稼働している個々の AI サービスやシステム自体を監査するケース(2.3.1)と、AI サービスやシステムを提供する組織で実施されている内部統制を監査するケース(2.3.2)に分類できる。前者はサービス、システム、モデルやプログラムそのものを対象とし、後者は組織内のガバナンス、マネジメントプロセス、規程類、人の作業、思考の連鎖等を対象として監査する。それぞれ監査の観点や手法、監査手続が大きく異なるため、どの部分を監査の検討対象とするかを事前に特定、合意せずに監査が実施された場合、関係者間の期待に沿わない監査結果となる。なお、この分類は便宜的なものであり、両者を組み合わせて実施することもある。

2.3.1 AI サービスやシステムの監査

AI サービスやシステム自体を監査するには、対象となる AI システムやサービスを特定する必要がある。AI システムの中核を担うのは学習済モデルだが、そのモデルが組み込まれた AI システムや、提供される AI サービスを監査する際には、学習済モデル以外の要素も監査対象になると考えられる。図 1 にて AI サービスやシステムについて監査の対象と考えられる構成要素ごとに分解した図を示す[17]。外枠から順に説明する。

---

[16]SOC2 報告書とは、米国公認会計士協会の定めるトラストサービス規準に従い受託会社が記述したセキュリティ等の主題に関連する内部統制に対して、監査人による手続実施結果と意見を表明した報告書である。その他、日本公認会計士協会「監査基準報告書 315 重要な虚偽表示リスクの識別と評価」(https://jicpa.or.jp/specialized_field/2-24-315-2-20230810.pdf)や経済産業省「システム監査基準」(https://www.meti.go.jp/policy/netsecurity/sys-kansa/sys-kansa-2023r.pdf)等で定められている基準やフレームワークも応用可能と考えられる。

[17]日本公認会計士協会「監査基準報告書 315 重要な虚偽表示リスクの識別と評価」(https://jicpa.or.jp/specialized_field/2-24-315-2-20230810.pdf)、IPA「共通フレーム 2013」、Jakob Mökander, Jonas Schuett, Hannah Rose Kirk, Luciano Floridi「Auditing large language models: a three-layered approach」等を参考にして作成。

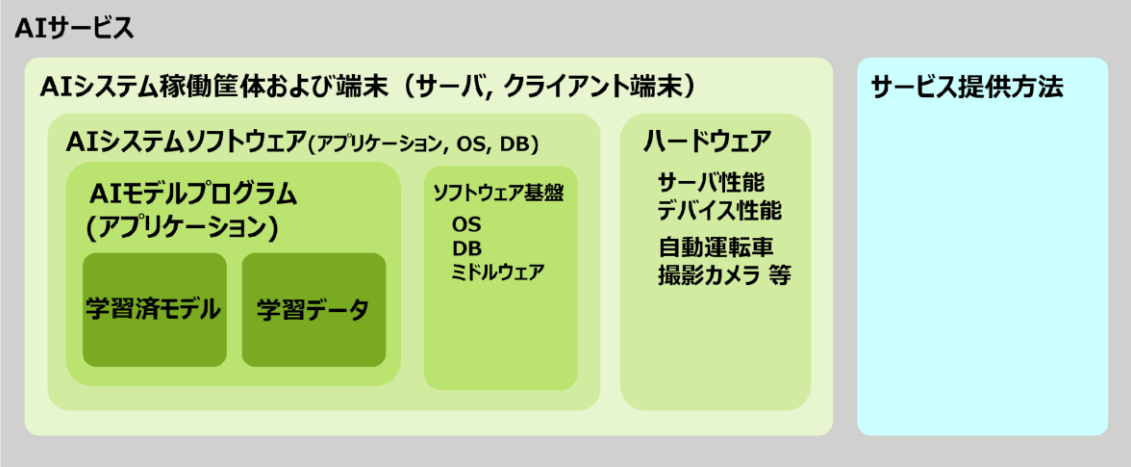

図 1 AI サービスやシステムの構成要素

(1) AI サービスの構成要素

AI サービスは学習済モデルが組み込まれた AI システムと組織がサービスを提供するフレームワークから構成される。組織がサービスを提供するフレームワークを検討する際には例えばサービスに AI が利用されていることを利用者に開示する方法が示されているか、サービスが利用する情報の取扱いについて適切に管理されているか等の論点が挙げられ、これも監査目的によっては監査対象となりうる[18]。

一方、AI システムは学習済モデルが格納されている AI システム稼働筐体や端末等のハードウェアを含めて監査対象となる[19]。この時、AI システムは、(i) 学習済モデルが格納されているサーバと、(ii)そのサーバに指示をしたり、AI の判断結果を表示したりするクライアントにわかれていることが多く、サーバ端末とクライアント端末の両方ともが監査目的によっては監査対象となりうる。

(2) AI システム稼働筐体や端末の構成要素

AI システム稼働筐体や端末は、AI モデルプログラムを含むソフトウェア以外にも、サービス提供に必要な AI システム稼働サーバやクライアント端末、デバイス等のハードウェアの性能も監査目的によっては監査対象となりうる。例えば、生体認証を行う AI システムの性能を監査する際、カメラ等のデバイスの性能も AI システム全体の性能に影響を及ぼす可能性があるため、これらのデバイスも監査対象として検討が必要と考えられる。またエッジコンピューティングを行っている場合は処理速度（レイテンシ）の観点からサーバの処理性能や設置場所、ソフトウェア基盤にクラウドサービスを利用している場合にはセキュリティや可用性の観点から提供サービスの設定、利用状況や利用リージョン（地域）等も

[18] IT サービスマネジメントの規格としては ISO/IEC 20000 IT サービスマネジメントシステムが挙げられる。

[19] 経済産業省「システム管理基準」（II.2.7）において、ハードウェアについても情報システムの構成要素として挙げられている。https://www.meti.go.jp/policy/netsecurity/sys-kansa/sys-kanri-2023.pdf

監査対象となりうる。

(3) AI システムソフトウェアの構成要素

AI システムソフトウェアのなかでも学習済モデルが格納されている AI モデルプログラム（アプリケーション）は監査対象として注目されるが、それ以外にもプログラムを稼働させるオペレーティングシステム（OS）やデータベース（DB）、ミドルウェア等のソフトウェア基盤も、従来のシステム監査と同様に監査対象となりうる。

(4) AI モデルプログラムの構成要素

AI モデルプログラムの監査には、アルゴリズム（計算・処理手順）の監査の研究がある（以下、アルゴリズム監査とする。アルゴリズム監査の系統的な文献レビューは(Bandy, 2021)を参照）。アルゴリズム監査とは、アルゴリズムの合法性、倫理性、安全性を評価、緩和し保証する研究と実践と定義できる(Koshiyama et al., 2022)。

また学習済モデルを作成するにあたっては、学習に利用される学習データも重要な検討要素であり、この学習データについても監査対象として挙げられる。データの収集課程や外れ値、偏りを考慮したデータの十分性、正解ラベルの設定方法なども監査項目となりうる(Batarseh et al., 2021)。

2.3.2 AI サービスやシステムを提供する組織で実施されている内部統制の監査

AI サービスやシステムを提供する組織で実施されている内部統制を監査するには、組織内のガバナンス、マネジメントプロセス、規程類、人の作業、思考の連鎖等に着目し、それらが適切に整備、運用されているかを対象として監査を実施する。ここでは、既存の内部統制監査と同様の枠組みにて監査を行う場合を想定して、その内容、手法、範囲と対象を整理する。

既存の COSO[20]や COBIT[21]等の枠組みに沿って概ね内部統制は構築され、監査も行われると想定されるものの、3 章に列挙した諸々の AI 監査を困難にする要因によって、従来のフレームワークだけでは対応できない論点もあると想定される。

(1) 監査内容

組織が実施する内部統制を対象とする既存の監査の例として、外部監査人による内部統制監査、企業内部の監査人による内部監査、マネジメントシステム監査のための指針

---

[20] トレッドウェイ委員会組織委員会（Committee of Sponsoring Organizations of the Treadway Commission）が提唱する内部統制フレームワーク。統制環境、リスク評価、統制活動、情報と伝達、モニタリングの要素で構成される。

[21] ISACA, ITGI が提唱する事業体全体を対象とした事業体の情報と技術のガバナンスとマネジメントのためのフレームワーク。Control Objectives for Information Technologies に由来。

（ISO19011）[22]に基づくマネジメントシステム監査等が挙げられる。何れの監査においても、組織において適切なルールが整備されており、組織内の人員がそのルールに則った適切な作業を行うことでルールが運用されているかを検討対象としており、プロセス、人の作業や思考の連鎖を対象とした監査である。これらの既存の監査の枠組みの中で AI サービスや AI システムが登場する場合においては、既存の監査の一環として AI サービスやシステムにかかわる組織、内部統制についても監査が必要になると考えられる。その他、IT(情報技術)を適切に管理する IT ガバナンスの延長として AI ガバナンス監査、AI サービスやシステムが個人情報やプライバシーの適切な管理に関わっているかを検討するのであれば AI コンプライアンス監査と言った形で、AI システムやサービスを提供する組織を切り出しての監査についても実施されうる。

(2) 監査手法

AI サービスやシステムを提供している組織の内部統制も、上述の COSO や COBIT 等の内部統制フレームワークに沿って行われることが想定される。内部統制を監査する場合には組織が直面しているリスクを低減するようなルールや、監査の立証命題を担保するようなルールが整備されており、組織内の人員がそのルールを適切に運用しているかについてヒアリングや関連証跡の閲覧により確認する手法が代表的である[23]。

(3) 監査範囲と対象

一般的に内部統制監査の対象となる個々の統制活動の範囲は、全社を対象とした広範な統制活動から個別のサービスやシステムを対象に実施される統制活動など幅広い。これは AI サービスやシステムに対しても同様であると想定される。

全社統制すなわち経営者レベルで管理すべき組織全体の統制活動を検討対象とする場合、社内 AI 利活用ガイドラインの制定や、AI 利用に関する教育研修体制等が監査対象として考えられる。一方、個別の AI サービスやシステム担当者レベルの統制活動を検討対象とする場合、AI サービスやシステムリリース（公開）前におけるテスト計画作成や、テスト実施結果の承認等の作業、手続等が具体的な監査対象の統制活動の例として挙げられる[24]。監査対象となるこれらの統制活動については従前の AI 以外のサービスやシステムを対象とした統制活動と類似するものと、AI サービス、システムに特徴的な統制活動の双方が含まれると考えられる。

---

[22] ISO190011 についての詳細は以下で公表されている。https://www.iso.org/standard/70017.html

[23] 金融庁「財務報告に係る内部統制の評価及び監査の基準」(https://www.fsa.go.jp/singi/singi_kigyou/kijun/20191206_naibutousei_kansa.pdf)等でより詳細な手法が紹介されている。

[24] 日本公認会計士協会「監査基準報告書 315 重要な虚偽表示リスクの識別と評価」(https://jicpa.or.jp/specialized_field/2-24-315-2-20230810.pdf)に一般的なシステムを対象としたシステム関連の内部統制が紹介されている。

2.4 AI 監査のタイミング

AI 監査のタイミングは、監査対象や種類（内部監査・外部監査）で異なる。本節では AI サービスやシステムのライフサイクルを分類したのち、監査対象や種類ごとに監査タイミングを整理する。

2.4.1 AI ライフサイクルの分類

OECD による AI ライフサイクルの定義[25]や一般的なシステム開発のライフサイクル[26]を参照とし、監査対象や種類ごとの議論しやすさを考慮に入れ、本稿は (1)新規開発、(2)機能変更や追加開発、(3)運用、(4)廃棄の 4 フェーズに AI のライフサイクルを分類した（図 2）。

図 2 AI ライフサイクル

(1) 新規開発

新規に AI サービスやシステムをリリースするには、企画段階でコンセプトを決定し、概念実証（PoC：Proof of Concept）を経て、開発可否が決定される。その後は開発、テストフェーズに進む。テストの結果、仕様書で定められた所定の品質に問題がないと確認されると、AI サービスやシステムがリリースされる。

(2) 機能変更や追加開発

稼働中の AI サービスやシステムの機能変更や追加機能の実装が必要となった場合、規模や要件に応じて再び企画フェーズからテストフェーズまでの必要な工程を経てリリースされる。データ変動に伴うモデルの精度劣化と、学習モデルを作成するアルゴリズムを変更することを前提に設計開発を行うため、機能変更や追加開発が頻繁に発生するのが AI サービスやシステムの特徴でもある[27]。

---

[25] AI サービスやシステムのライフサイクルとして OECD は(1) 設計・データ・モデリング、(2) 検証と妥当性確認、(3) 実際の運用環境への実装、(4) 運用と監視の 4 つのフェーズに分類している。OECD. "Scoping the OECD AI principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO)", OECD Digital Economy Papers, No. 291, OECD Publishing, Paris, https://doi.org/10.1787/d62f618a-en

[26] IPA「共通フレーム 2013」では利害関係者要求定義プロセスからソフトウェア処分プロセスまで、より詳細なプロセスが紹介されている。

[27] 秋元一郎ほか「AI ビジネス大全」（プレジデント社）

(3) 運用

稼働中の AI サービスやシステムに対しては、バッチジョブの実行やエラーモニタリング、障害対応等が必要となる。この運用フェーズには様々な定義が考えられ、監視・保守対応だけではなく、(2)の機能変更や追加開発も運用フェーズと見なすこともできる。

(4) 廃棄

AI サービスやシステムは終了に向けた収束作業・手続を経たのち、最終的に廃棄のフェーズとなり、保有データやプログラムの適切な処分、利用者への周知が行われる。

2.4.2 対象ごとの監査タイミングの分類

AI 監査の対象は 2.3 と同様、AI サービスやシステムを対象とした場合と、AI サービスやシステムを提供する組織、内部統制を対象とした場合に分類できる。

(1) AI サービスやシステムの監査タイミング

従来のシステム監査、特に外部監査においては新規開発のリリース後に本番稼働している対象システムの監査を実施することが多い。しかし大量のデータからパターンや傾向を読み解く帰納法的な AI サービスやシステムを監査する場合、AI サービス自体の妥当性、開発要否の判断の妥当性を判断する必要が出てくるため、必要に応じてテスト、開発フェーズ、PoC や企画フェーズも含めてリリース以前のフェーズに遡った監査を実施することも想定される。

また、特に継続学習を実施している AI サービスやシステムの場合、監査手続を実施した時点での AI 出力結果の精度と、その監査結果を利用する時点での AI 出力結果の精度に差異が発生する可能性については留意する必要がある。例えば、同じ入力データを対象とした場合であっても、追加学習の前後で出力結果が変更となるケースがあるため、その場合には正解率や適合率といった精度指標が時点によって異なる結果となる。

(2) AI システムやサービスを提供する組織で実施されている内部統制の監査タイミング

AI システムやサービスを提供する組織で実施されている内部統制を対象とした監査の場合、企画から廃棄までのライフサイクル全てが監査対象となると想定される。

AI 監査実施のタイミングとしては、外部監査の場合は運用初年度の対象サービスリリース後と考えられるが、継続監査の場合や複数サービスを対象とする場合には、年間を通じての評価が想定される。なお、期間については実務上その他の監査と整合させることが合理的であり、年単位の区切り（通常 1 年）が一般的と考えられる。しかし AI は技術進化が早く、また対象とする AI のリスク度合いによっては、これに限らない適切な頻度で実施される可能性がある。

2.4.3 種類ごとの監査タイミングの分類

監査を行う主体により内部監査と外部監査に分類され、その監査タイミングは異なる。

(1) 内部監査

同一組織内で行われる内部監査の場合、任意に監査目的や監査対象が設定できる。そのため、ライフサイクル全てのフェーズを監査対象とすることも、特定のフェーズを深堀して監査を実施することもできる。例えばAIサービスやシステムの品質やガバナンスの向上、今後監査の実施が容易となるようなシステム設計を見込んで、企画フェーズから内部監査人が関連部署と伴走して監査を実施することも可能である。

(2) 外部監査

従来のシステム監査では、開発フェーズ以降を中心に外部監査が実施される[28]。ただし、AI サービス提供にあたってはサービスそのものの妥当性、開発要否の判断の妥当性も重要な論点であるため、必要に応じて企画や PoC フェーズも含めた監査が外部監査においても検討される可能性がある。

2.5 AI監査の実施者要件

AI 監査の実施者要件を(1)専門性要件、(2)独立性要件、(3)組織要件、(4)監査人の法的責任の 4 つの観点から整理する[29]。このうち、(1)専門性要件と(2)独立性要件は内部監査と外部監査で共通であるが、外部監査はこれに加えて(3)組織要件や(4)監査人の法的責任等の観点が追加される。

2.5.1 専門性要件

AI 監査の実施には幅広いスキルや経験による多様な専門性が求められる。監査論の理解、被監査企業や提供しているサービスに関連する業界知識、AI に限らない IT 領域の知識や経験に加えて、AI 固有の技術的知見、倫理や文化、法規制等の非常に幅広い範囲での知識や経験が必要となる。このように、AI 監査の実施者に求められる専門性要件は非常に高く、実施の際に求められるスキルや経験を個人で全てカバーすることは極めて困難である。従って、現実的には複数人やチームによる監査が必要となると考えられる。また、AI サービスやシステム以外も対象とした包括的な監査を実施する場合、監査制度の設計によってはAI監査の実施者を専門家と位置付けた上で、AIサービスやシステムに関連する部分では専

---

[28] ただし、品質マネジメントシステム（QMS）監査など、システム監査以外の場合は、必ずしもこの限りではない。

[29] 金融庁企業会計審議会「監査基準」(https://www.fsa.go.jp/singi/singi_kigyou/kijun/20201106_kansa.pdf)および日本内部監査協会「内部監査基準」(https://www.iiajapan.com/leg/pdf/guide/20140601_2.pdf)にて要求されている事項を整理した。

門家利用というスキームを用いて監査が実施される可能性もある[30]。ただし、監査チームの規模については AI 監査に投入可能なコストも加味した上で最終的には決定されるものと考えられる。

2.5.2 独立性要件

従来の監査と同様、AI 監査は被監査企業、部門との利害関係のない独立性を担保したメンバーや組織による実施が求められる。独立性要件は外部監査、内部監査それぞれにおいて定められている[31]。

2.5.3 組織要件

AI 監査の実施結果に対する信頼性を確保するため、監査の実施組織は品質や独立性等の観点で一定の基準を満たす組織要件が必要となる。組織として監査品質や AI 監査実施者や実施組織の独立性を確保するための体制やモニタリングを実施することで、AI 利用者が信頼して監査結果を利用できる。

上記のような組織要件を満たした組織を存在させるためには、組織に対する認定・認定制度やモニタリングを実施する機関や役割についても検討しなければならない。また、このような組織要件を満たしていない監査実施機関が実施した監査結果については、実態が正しく反映されていない可能性があることにも留意が必要である。

2.5.4 監査人の法的責任

AI 監査実施者が正当な注意[32]を払って監査を実施したものの、監査結果が誤っていた場合、AI 監査実施者が負うべき法的な責任範囲を検討する必要がある[33]。例えば、自動運転車に搭載されている AI システムの安全性について監査を実施し適正意見を表明したにも関

---

[30] 外部監査における専門家利用は日本公認会計士協会「監査基準報告書 620 専門家の業務の利用」(https://jicpa.or.jp/specialized_field/2-24-620-2-20230810.pdf)により定められている。

[31] 外部監査においては日本公認会計士協会が定める「倫理規則」(https://jicpa.or.jp/specialized_field/files/2-22-0-2-20221031.pdf)において精神的独立性と外見的独立性の双方が公認会計士に要求されている(セクション 120.15 A1）。また、内部監査においても、独立性と客観性として内部監査部門は組織上独立していなければならず、かつまた内部監査人は内部監査の業務（work）の遂行にあたって客観的でなければならないと定められている(IPPF1100, IIA(2017d))。

[32] 日本公認会計士協会が定める「倫理規則」(https://jicpa.or.jp/specialized_field/files/2-22-0-2-20221031.pdf)において、「職業的専門家としての能力及び正当な注意の原則は、会員に対し、適切な専門業務を提供できるよう、専門知識及び技能を必要とされる水準に保持するとともに、適用される基準及び法令等を遵守し、職業的専門家としての正当な注意を払うことを求めている。（後略）」とされている（セクション 120.16 A2(3)）

[33] 日本公認会計士協会法規委員会研究報告第 1 号「公認会計士等の法的責任について」(https://jicpa.or.jp/specialized_field/files/2-15-1-2-20160801.pdf)においては、民事、行政、刑事、その他の 4 つの区分にて公認会計士等の責任について説明されている。

わらず、該当の自動運転車が AI システム機能に起因して事故を起こした場合等は、監査人の責任範囲について慎重な検討が必要なケースと想定される。

AI 監査実施者の責任が重すぎると監査の担い手がいない、あるいは不足することも考えられる。そのため、AI 監査実施者を守る免責要件や保険制度の必要性等も検討していく必要がある。

2.6 AI 監査の関係者と関係組織

AI サービスやシステムの監査に関わるのは監査人と被監査企業だけではない。図 3 に現時点で想定される AI 監査に関係する関係者と関係組織を整理した。以下、(1) AI サービス提供組織、(2)AI サービス利用者、利用組織、(3)外部監査実施者、(4)標準化団体/認証・認定機関、(5)公的機関、(6)民間団体と(7)その他関係者を説明する。

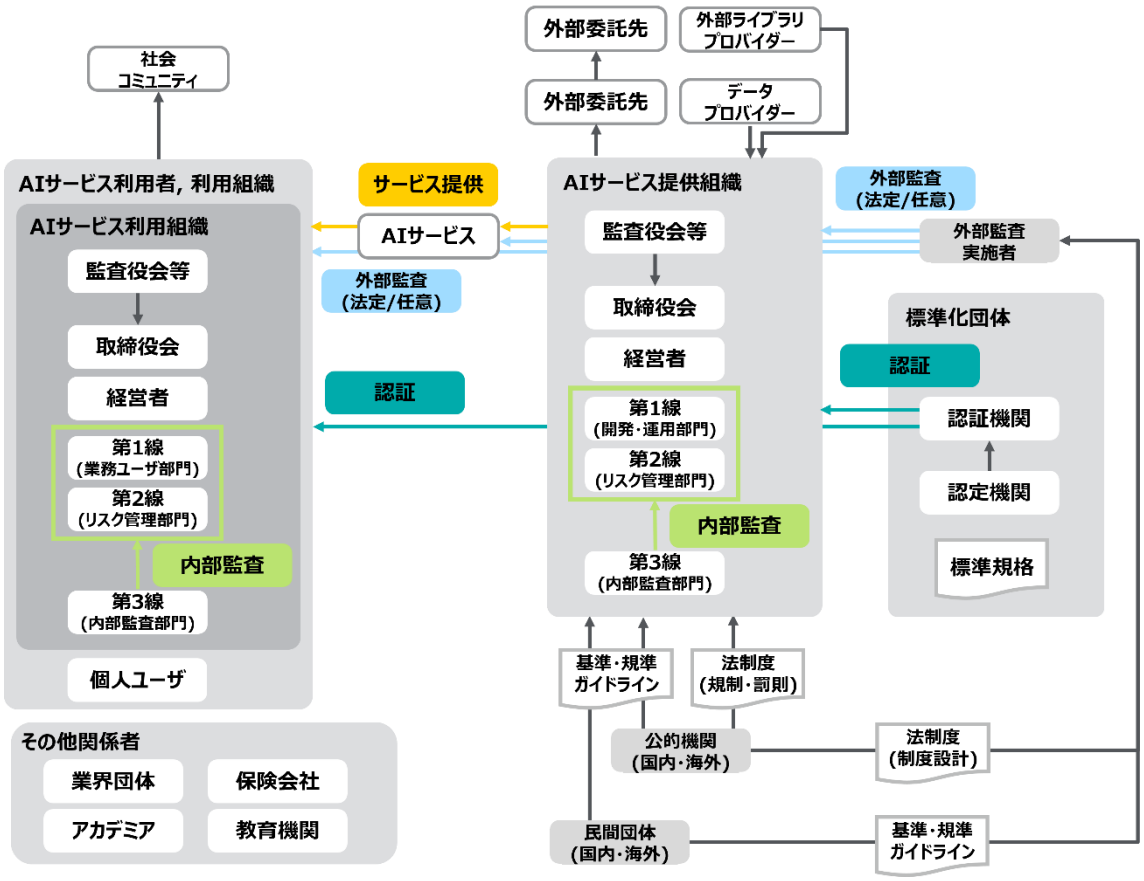

図 3 AI 監査の関係者と関係組織

2.6.1 AI サービス提供組織

AI サービスやシステムは一つの組織で提供される場合と、組織をまたいで提供される場合が想定される。

(1) 組織内ガバナンス

AI サービスやシステムを提供している組織は、2.3 の AI 監査の対象で分類したように、サービスやシステムそのものの監査、あるいは組織が実施する内部統制の監査を受ける。

組織としてリスクの管理を含む適切なガバナンスを行うためのモデルとして内部監査人協会が提唱する 3 ラインモデルがある[34]。AI サービスやシステムを提供している組織の 3 ラインは、AI サービスやシステムの開発や運用を実施する部門が第 1 ライン、それを牽制するリスク管理部門が第 2 ラインである。経営者及び執行部門は第 1 と第 2 ラインの役割としてガバナンス強化、マネジメント向上を目標として掲げる。これに対し組織内で独立した立場から経営者及び執行部門のマネジメントを検証、内部監査する部門が第 3 ラインである[35]。この 3 ラインモデルが適切に機能しているかを評価することも監査役や監査等委員会による監査(必要に応じ外部監査を利用)の観点として想定される。

(2) 組織をまたぐガバナンス

サプライチェーンが複雑、高度化する中、AI サービスやシステムを自社のみで構築することは実務上困難であり、外部組織の協力を得てサービスやシステムが構築される場合も多い。この場合、システム構築をサポートしている外部委託先や外部ライブラリ、データプロバイダーも被監査対象として含めるか、その範囲も含めて検討する必要がある。また AI サービスやシステムの提供組織と外部委託先との関係性は多様であり、例えば以下のケースでは AI サービスの利用者や主な開発者が異なるため、それぞれ異なる観点での監査検討が必要と考えられる[36]。

A) AI サービス、システムを自社及び外部協力を得つつ構築しており、外部利用者向けに AI サービスを展開している企業
B) AI サービス、システムを外部に委託して構築しており、構築した AI システムについては自社のみで利用している企業
C) 他社向けの AI システム構築や、パッケージの販売を行うソフトウェアベンダー企業

2.6.2 AI サービス利用者, 利用組織

AI サービスやシステムを利用している個人や組織も、AI サービス提供企業と同様に利用者の立場として監査を受ける可能性がある。この場合、AI サービスをどのように利用して

---

[34] 3 ラインモデルは、2020 年に組織のリスク管理・統治活動のモデルとして内部監査人協会（The Institute of Internal Auditors: IIA）が公開したものである（IIA(2020)）。

[35] なお三様監査の観点からは、内部監査の他に監査役監査も考えられるが、本稿では詳細な検討を割愛する。

[36] 経済産業省「AI・データの利用に関する契約ガイドライン」(https://www.meti.go.jp/policy/mono_info_service/connected_industries/sharing_and_utilization/20180615001-1.pdf)を参考とし、代表的と考えられるケースを想定、提示している。

いるか、AI が判断した結果をどのように組織内で活用しているかが主な監査の観点となる。そのため、3 ラインモデルのうち第 1 ラインは AI サービスを利用する業務ユーザ部門となる。

なお自社で AI サービスを構築し、自社でそのサービスを利用している場合においては、2.6.1 のサービス提供企業としての観点と、2.6.2 の AI サービス利用者としての双方にて監査を実施する必要がある。

2.6.3 外部監査実施者

被監査企業に対して外部監査を実施する独立した第三者は、法令等に基づき、完全に独立した立場より検証を実施し、監査結果について監査意見として公表する。外部監査の監査人の実施者要件として検討すべき論点として、専門性、独立性要件、組織要件、監査人の法的責任等がある（2.5 参照）。

2.6.4 標準化団体/認証・認定機関

内部監査や外部監査の指標となる規準や規格を策定する標準化団体の例には、国際標準化機構（ISO）、国際電気標準会議（IEC）、米国電気電子学会（IEEE）等に代表される国際標準化団体の他、国内では日本産業標準調査会等がある。これらの団体が定めた基準や規格が監査では利用あるいは参考にされる。ISO 規格等に合致しているかを審査する第三者を認証機関といい、その認証機関を審査する認定機関（日本では JAB）がある。

2.6.5 公的機関

監査を制度として整備していく方法の一つとして法制度による整備がある。法制度は立法者により制定されるため、立法者を含む国内外の公的機関も AI 監査の関係者として考慮に入れる必要がある。

一方、AI 監査を実施する際の具体的な手続や、監査結果の判断、AI サービス提供組織側での規程、手続類については実務上、公開されている各種基準等に従うことが想定される。これらも公的機関により整備される場合があり、例えばシステム監査の国内基準として 2023 年 4 月に改訂された経済産業省「システム監査基準」[37]、「システム管理基準」[38]がある。

2.6.6 民間団体

AI サービス提供組織や AI 監査実施者が遵守すべき各種基準には民間団体が発行しているものも多い。例えば、財務諸表監査においては、被監査企業が遵守すべき会計基準は民

[37] 経済産業省「システム監査基準」、https://www.meti.go.jp/policy/netsecurity/sys-kansa/sys-kansa-2023r.pdf

[38] 経済産業省「システム管理基準」、https://www.meti.go.jp/policy/netsecurity/sys-kansa/sys-kanri-2023.pdf

間の団体である企業会計基準委員会（ASBJ）が開発・公表している。従ってこれらの基準を発行する国内外の民間団体も AI 監査の関係者として考慮する必要がある。

2.6.7 その他関係者

その他の関係者として、継続的な技術革新及び利用者教育という点からはアカデミアや教育機関、AI サービスを提供する個々の産業特有の規制や協調という点からは業界団体、監査とは異なる観点でのリスク対策として保険制度を定める点からは保険会社等、多くの関係者が AI 監査に直接的、間接的に関係しているため AI 監査の関係者として考慮する必要がある。

3. AI 監査を困難にする要因

AI サービスやシステムの監査を困難にしている概念的、技術的、制度的、社会的な要因が指摘されている（Mökander et al.(2023)）。本稿では(1)AI 技術の複雑性、(2)AI 監査制度設計の未整備、(3)AI 監査の実施基準設定の難しさ、(4)被監査対象者の範囲に起因する複雑性、(5)AI 監査に対する需要と供給のアンバランスの 5 項目に焦点を当てて説明する。

3.1 AI 技術の複雑性

従来のシステム監査では、業務要件で定められた正解通りにシステムが設計、プログラミング等により実装されており、入力データと構築されたロジックにより一意に想定された通りの値がシステムより出力されるかと言った観点及び検証手法により、精度等の評価が実施されていた。しかし AI システムは従来の IT システムとは異なり、事前に定義された一意の値を出力する想定で構築されていない。

また AI システムの構築においては、判断のロジック部分がブラックボックス化されていることが多く、該当のロジック部分の検証が困難である。仮にプログラムのソースコードや特徴量等のパラメータ設定が公開され、数式や条件判定文の形で該当ロジックを再現することができたとしても、数式やパラメータは大量かつ複雑であることが想定されるため、そのロジック自体の妥当性に関する完全な判断は難しい。

さらに AI システムの構築方法によっては、学習段階でランダム要素を含めてモデル学習させている場合もある。このような場合、仮に同じ方法で学習を実施しても、異なる学習済モデルが作成されることが想定される。そのため、このようなランダム要素まで含めて監査上の判断をしなければならない。

加えて AI 監査のタイミング（2.4）で言及した通り、継続学習を実施する AI サービスやシステムもある。この場合、監査実施時点での評価と監査結果利用時の評価が異なる場合があり、AI 監査の実施を難しくする。継続学習、あるいはバグで出力が変わったのかの見極めも難しく、監査の再現性という観点からも AI 監査の難度が上がる。

最後に、人と機械の関係性の複雑化も AI 監査を困難にする。AI サービスやシステムが人の意思決定の支援として用いられるような設計にもできれば、人の介在なく意思決定を

行う設計にもできる。医療行為のようにAIシステムは診断の補助として用いて最終的な判断は医師が行うと定められている領域もあり、技術がどのような前提条件、文脈で使われているかによって監査の位置づけも変化する。

3.2 AI監査制度設計の未整備

2023年9月現在、AIに特化した監査実施者の要件や品質マネジメント体制等の一般基準は制定されていない。また監査手続を設計する際にも標準化された基準、規準（実施基準）が確立していないため、監査人が独自で手続を設計し、監査結果を判断する必要がある（阿子島・福田　2020a, 2020b, 2020c）。

加えて、国内・国際ルール、経営者、監査人、認証機関などの関係者が順守すべき統一的なルール形成、確固たる合意がない。そのため、それぞれの関係者が何を遵守すべきかの理解が困難となってしまっている。このようにAI監査に対する監査制度設計が未成熟であることが、AI監査の実施を困難にする一つの要因となっている。

3.3 AI監査の実施基準設定の困難性

監査人が独自で手続を設計する場合でも、AI監査の実施基準を設定することは困難である。例えばAI監査には様々な立証命題(2.2)が想定されている。その中には、公平性のように、一意の定義が難しく、判断や検証が困難なものも含まれている。例えば、技術的な公平性の指標は大きく2つあるが[39]、これらは同時に達成することができないとされている。

さらに異なる価値間のトレードオフ関係も考慮に入れなければならない。一般的にAIの判断精度の確保と説明可能性はトレードオフの関係にある。また、監査の立証命題間でもトレードオフとなる項目が想定される。例えば、安全性や正確性の観点での監査と、効率性の観点での監査では検討すべき監査上の論点や検証項目が異なるため、一度に複数の観点から監査を実施することは困難である。

その他、AI監査実施の際には学習データに関してデータガバナンスの論点も併せて検討を実施し、監査を実施する必要がある。しかし学習データの偏りや学習データの十分性に関する指標の設定が困難であることもAI監査の実施基準の観点から監査実施を困難にする。

3.4 被監査対象の範囲に起因する複雑性

AI監査の関係者と関係組織で紹介したように(2.6)、AIサービスやシステムには利害関係者が多く、どの範囲までを監査対象に含めるべきかの線引きが困難である。

---

[39] 一つは「一人の個人が集団属性に関係なく、他の個人と同様に扱われている状態を指すindividual fairness」であり、もう一つは「集団の中で男女等のセンシティブグループ間の公平性を指すgroup fairness」である。このgroup fairnessの主要な規準としてstatistical parity（Dwork 2012）やequalized odds（Hardt 2016）がある。

3.4.1 組織を超えた AI システムの開発と運用

AI システムやサービスの特徴として、国や組織を超えて開発、利用されることが挙げられる。単一の組織内に留まらず、外部委託先等の複数組織を監査対象として識別し、組織を超えて監査を実施する必要が出てきた場合、AI 監査の実施が困難となる。

例えばシステム開発や運用に外部ベンダーが関わっている場合、どの範囲までソフトウェアベンダーを監査の対象として含めるべきか検討が必要である。開発や運用において内部統制の実施主体がベンダーであり、委託元が関与していない場合は、ベンダーを外部委託者として取り扱った上で、ベンダー側にて実施されている内部統制を監査対象に含めるべきと考えられる。

一方で、外部委託者に対する監査を実施する場合、対象となる外部ベンダーに対してどの範囲までの監査が実施可能かという課題もある。またベンダーがさらに他のベンダーに再委託している場合も存在する。実務的には多重委託が行われている場合、締結されている契約の観点から複数の委託先を監査することが困難であることが予想される。仮に監査が実施できたとしても、委託先の監査費用を誰が負担するのかと言った金銭面での問題も想定される。

3.4.2 外部ライブラリの利用

AI システムの構築においては外部ライブラリとして一般に公開されている機能を採用するケースが多い。このような場合、AI システムが正しく機能しているという点は外部ライブラリで提供されている機能や処理が正しく機能していることに依拠しているとも考えられる。これらのライブラリで提供されている機能が正しく機能しているという点を所与とせずに、外部ライブラリ機能やその提供者を監査対象に含めると、さらに監査範囲が広範になる。

3.4.3 学習データのガバナンス

学習データの管理についても、データの収集業者や公的データの提供者、学習のための正解ラベルの設定者等、多くのステークホルダーが関与している。学習データの妥当性については AI 監査を検討する中で不可欠であると考えられるが、学習データの関係者やそのガバナンスまで監査対象を含める場合には、さらに監査範囲が広範になる。

3.5 AI 監査に対する需要と供給のアンバランス

AI 監査の必要性（1.1）で整理したように、AI サービスやシステムに対する監査の必要性は AI サービスやシステムが社会に普及していくにつれて高まっている。しかし、AI 監査にて保証される内容に対する期待値の不一致や人材の不足、インセンティブの欠如等により、AI 監査を適切な対象や実施者要件のもとに実施することが困難である。

3.5.1 AI 監査で保証される内容に対する期待の不一致

AI 監査の対象は、2.3 で整理したように AI システムやサービスそのものと、それを提供する組織で実施されている内部統制の二つに分類される。このうち 3 章で今まで紹介してきた技術、制度あるいは社会的な要因により AI サービスやシステムそのものの監査は困難である場合も想定される。そのようなケースで AI 監査を実施する場合には、AI サービスやシステムを提供する組織で実施されている内部統制が適切に機能しているかを保証することのみを保証する形となる。さらに、監査の対象が限定的であることに加えて、その保証内容についても様々な制約や限界がある[40]。これに対し社会的には「AI サービスが 100%正しく動作し、安全であること」など AI システムやサービスの保証が期待されており、またその期待についても AI サービスやシステムの特性が考慮されていない場合、AI 監査の要求事項と実施内容の間に期待ギャップが生じる。

3.5.2 AI 監査を実施する人材の不足

AI 監査の実施には、実施者要件(2.5)で整理したように、監査に関する専門知、AI に関する技術的な知識や法的・倫理的な内容をカバーする幅広い知見が要求される。しかし現時点ではそのような AI 監査を実施できる人材は不足していることが想定される。

3.5.3 AI 監査実施者の需要と供給の不一致

監査結果が誤っていた場合、監査実施者は法的な責任を負うことも想定される。しかし 3 章で紹介してきたように技術的、制度的、社会的な要因により監査の実施が困難であるため、監査人の責任と監査報酬のバランスが取れなくなることも考えられる。そのため、AI 監査の需要が高まったとしても外部監査人にとって監査を実施するインセンティブが乏しく監査の実施が困難になる。

3.5.4 被監査企業が AI 監査を受けるインセンティブの欠如

監査実施主体に監査を実施するインセンティブが乏しいだけではなく、被監査企業側からの AI 監査に対する需要もまだ少ない。2023 年 9 月現在、本研究会で調べた限りでは AI 監査を必須とするとするような法的拘束力を有する規制や罰則等は日本国内にはなく、海外においても州や条例を除いた国レベルのものは存在しない[41]。また、3 章で整理してきた通り、AI 監査の実施は困難かつテーマが多岐にわたるため、高品質な監査を実施する場合には多大なコストがかかり、これも被監査企業側が監査を受けるインセンティブが働かない要因となる。

---

[40]「AI 監査」そのものではないが、例えば「会計監査」においてはその機能的限界として(1)会計判断の相対性(2)取引実態や事象や事実への監査人関与の限界(3)契約事項としての監査の限界があげられている(山浦(2008),p.12)。

[41] 前述のように米国ニューヨーク市が、AI を活用した採用を規制する法律を 2023 年 7 月に施行した。

4. AI 監査の想定事例：採用 AI

採用 AI という具体事例を用いて、2 章の AI 監査をめぐる論点と 3 章の AI 監査を困難にする要因を紹介する。ここで事例として挙げる採用 AI とは、求職者の応募書類と面接の録画データをもとに採否判断を決定するサービスを想定する。なお、通常のシステム監査において検討すべき論点と重複するような論点も想定されるが、ここでは AI サービス、システムにそれらの論点を取り込みつつ、AI 監査に特徴的な論点を中心に取り上げていく。

4.1 AI 監査の必要性

AI サービスを利用する企業としても、AI の判断結果の精度に関してだけではなく、公平性やプライバシー等の観点が考慮されている等が監査で保証されていることが、被採用者の安心と企業の評判にもつながる。また採用 AI は欧州 AI 規制法案においてもハイリスク AI と位置付けられており、AI サービス提供者は事前に適合性評価手続を経る必要があり、法案が施行されれば監査の必要性が法的にも求められることとなる。このような点からも AI サービスやシステムの提供側と利用者側の両方に監査を受けるインセンティブが発生する。

4.2 AI 監査の立証命題

AI 監査の必要性で述べられている通り、AI の判断結果の精度のほか、公平性、プライバシー等を立証命題として監査が実施される可能性がある。実際には何が立証命題となるかは、被監査企業との議論だけではなく、社会的な要請も関わってくる。

4.3 AI 監査の対象

AI システムサービス自体を監査する場合と、AI サービスを提供する組織で実施されている内部統制を監査するケース双方を考える。

4.3.1 AI システムやサービス自体を監査するケース

図 1 で挙げた項目を具体的に検討していくと、例えば以下のような項目が監査対象となると考えられる。

(1) 学習済モデル

実際にプログラムとして完成されたロードモジュールファイルやソースコード、特徴量を含む各種パラメータファイル等。その他モデルの仕様を定義した設計ドキュメント等も監査対象になると考えられる。

(2) 学習データ

過去の社内の人事データや、公表されている統計データ等が監査対象になると考えられる。

(3) ソフトウェア基盤

AI モデルやプログラムを実際に稼働させているサーバのオペレーティングシステムやデータベース等が監査対象になると考えられる。加えて、オンプレミスでサーバを準備するのではなく、クラウド上にサーバを構築している場合にはクラウド事業者が提供している各種ソフトウェアサービスの利用状況も監査対象になると考えられる。

(4) ハードウェア

AI システムへの入力データとなる面接を録画するカメラ機器や録音機器等の性能についても監査対象になると考えられる。また、AI モデルを稼働させているサーバの処理能力や冗長性、クラウドサービスを利用しているのであれば、ソフトウェア基盤同様にバックアップサービス等の利用状況や利用リージョン（地域）等も監査対象になると考えられる。

(5) その他サービス提供方法

採用 AI を利用して採否判断やスコアリングを実施していることが求職者に対して開示されているか、どのように第三者チェックがされているか等が監査対象になると考えられる。

4.3.2 AI サービスを提供する組織や内部統制を監査するケース

(1)経営層、(2)利用部門である人事部、(3)システム開発や運用部門であるシステム部にて実施されている内部統制を対象に監査が実施されると考えられる。ただし以下の内部統制はあくまで考慮すべき一例であり、実際には被監査企業側での目的思考、業務プロセスやシステム開発、運用プロセス、AI サービスの利用方法を考慮した上で、検討すべき内部統制は広範囲となる。

(1) 経営層

経営層の AI サービスの開発・運用方針を確認するために AI 活用ポリシーや教育体制等が経営者により確立、運用され、PDCA サイクルが有効に機能しているかを確かめること等が考えられる。

(2) 利用部門である人事部

AI の判断結果通りに実際に採否を決定することや、AI にて判断不能と判断された場合には面接の担当者が手動にて判断を実施することを規定するルールが制定され、正しく運用されているかを確かめること等が考えられる。また、学習データを人事部が準備している場合には、正解ラベルの設定に関する判断基準が定められており、設定結果について複数人での確認が実施されているかを確かめること等も対象の内部統制の検証手続として想定される。

(3) 開発、運用部門であるシステム部

各種設計やテストの計画、実施結果について確認や承認ルールが定められており、その通りに運用されているか、リリースを許可する精度の閾値が定められている場合には、その閾値を超えたプログラムのみがリリースされていることを確かめること等が考えられる。

4.4 AI 監査を困難にする要因

AI 監査の対象となる AI サービス、システムを特定して具体的な議論する場合、その業界や利用される文脈に関する論点も考慮に入れる必要があり、一般論で議論しているより AI 監査がより複雑になる。例えば採用 AI の場合には雇用慣行や教育に関する社会制度等の追加的な論点についても検討対象となるが、これが議論をより複雑にしている。

4.4.1 AI 技術の複雑性

監査実施時点では適切な採否判断を出力していても、その後の追加学習や事業領域や社会制度の変更により採否の妥当性が監査結果利用時点では異なる可能性がある。また採否判断のロジック部分を数式や条件判定として表現することが困難であり、仮にロジックが明示的に確認できた場合であっても、そのロジックが妥当なのかどうかの判断は極めて困難であることが想定される。

また人事部における採否判断について、採用 AI の判断結果に完全に依拠するのかに関しても前提条件を考慮する必要がある。AI の判断結果は補助的に利用するのみで、最終的には人が判断しているのであれば、採用 AI の監査への位置付けも変わる。

4.4.2 AI 監査制度設計の未整備

欧州 AI 規制法案が施行された場合、採用 AI に関する監査が必要とされることが想定されるが、現段階においては AI 監査実施者の要件や品質マネジメント体制等の確固たる一般基準は確立していない。2023 年 7 月にはニューヨーク市が AI を活用した採用などを規制する法律を施行したため、今後、この法律が一つのベンチマークとなることが期待される。

4.4.3 AI 監査の実施基準設定が困難

AI 監査の実施基準を考慮するにあたり、精度や性能を評価するための指標を検討する必要があるが、例えば優秀かつ退職しない人材を正確に判断できるかどうかの精度を確かめる指標には、バイアスの観点は重視されていないことも想定される。仮に優秀な人材を正確に見抜くモデルであっても、差別的な判断をしているケースも考えられる。

また何をもってバイアスがないか、何をもって公平な結果と判断をするかの基準作りは難しく、完全な社会的な合意には至りにくいとも考えられる。そもそも AI サービスを利用しない採用過程においても、公正な採用選考の考え方は示されているものの、雇用主にも採用方針や採用基準、採否の決定など採用の自由が認められており、公平な採用が行われているかの画一的な基準は存在しない。

さらに採用者はAIサービスの判断結果が正しかったのか事後的に検討することが可能だが、不採用者が真に不適切な人材だったかは事後検討できないため、精度の評価は一部のケースのみに限定されることになる。

4.4.4 被監査対象者の範囲に起因する複雑性

採用AIサービスを一社内で構築しているか、組織を超えて開発や運用をしているかによって、論点は異なる。本事例では、組織をまたがって開発、運用をしていると仮定する。

(1) 組織を超えたAIシステムの開発と運用

AI システムの構築にあたり、実際に開発を担当しているソフトウェアベンダー企業、その企業がさらに開発を委託している開発ベンダー企業等が存在する場合、どの範囲までを監査対象として含めるのかの判断が困難である。

(2) 外部ライブラリの利用

ディープラーニングや勾配ブースティング、ランダムフォレスト等の手法で機械学習を実現し、モデル作成の元となっている各種外部ライブラリについての監査要否を検討する必要がある。また、AI システム構築サービスを利用してモデルを構築した場合には、利用したAIシステム構築サービスに対しても監査要否の検討が想定される。これらの外部ライブラリや構築サービスについて実際に監査を行うには様々な障壁も想定され、どの範囲までを監査の対象に含めるかの検討は困難であることが想定される。

(3) 学習データの管理

民間の人材紹介会社からデータ提供を受けている場合、そのデータ管理の監査要否を検討する必要があるが、仮に監査が必要となった場合でも実際に監査を実施するには様々な障壁が想定される。また公的機関が公表している統計情報についても、利活用するコンテキストの違いに関わらず無条件で監査不要として利用して良いのか等、公的機関による公表データの取扱いについても判断が困難である。

さらに学習元となる過去の人事データに対して、正解ラベルの設定を人事部が実施している場合、人事部のラベル設定作業の妥当性も含めて監査が必要となるため、監査の範囲が広範となることが考えられる。

4.4.5 AI監査に対するニーズと供給のバランス

採用 AI はすでに国内外において、一定程度使われている。今後、採用 AI を含むハイリスク AI の監査の必要性が高まる可能性がある一方、AI 監査にて保証される内容に対する期待値の不一致や人材の不足、インセンティブの欠如等により、AI 監査を適切な対象や実施者要件のもと実施することが困難となる可能性もある。

(1) AI 監査で保証される内容に対する期待の不一致

世間一般からの AI 監査に対する期待として、例えば「採否判断結果に対して 100%の正解を誇る AI」や「万人が考えるバイアスを完全に排除した AI」かどうかを監査するというニーズが想定される。一方で、実際に監査で確認、保証可能な内容の限界としては、「採否に関する AI の判断結果の精度として、各評価指標が事前に定義された閾値に収まっているか」の確認や、「監査人や経営者が定義した公平性に沿っているか否か」の確認に留まることも想定される。このような場合には AI 監査の内容について期待ギャップが生じることとなる。

(2) AI 監査を実施する人材の不足

採用 AI サービスの監査を実施するには、監査に関するスキルや AI に関する技術的な知見以外に、雇用や人事をめぐる社会的、法的、倫理的な観点もカバーする幅広い知見が要求される。特に、現在 AI 採用サービスに関する監査は監査人が独自で手続を設計し、監査結果を判断する必要があり、これらを実現する知見、経験を持つ人材の不足が考えられる。

(3) AI 監査実施者の需要と供給の不一致

求職者側から不当な雇用差別に対する賠償請求をされた場合、採用 AI サービスを開発、利用している経営者だけでなく、AI サービスを監査した外部監査人に対する賠償責任の有無が議論になる可能性がある。賠償責任が発生する際、その責任が監査報酬と見合わない場合、監査実施者が不足することも考えられる。

(4) 被監査企業の AI 監査を受けるインセンティブの欠如

現状国内において採用 AI の監査を義務付ける法令は存在しないため、利用企業側で監査を受けるインセンティブが乏しい。また対象の検証領域が多岐にわたる場合、本格的な監査実施には多額の監査報酬も想定されるため、採用 AI サービスの利用による便益が外部監査の報酬に見合わないとして、採用 AI の利用を取りやめる企業が登場することも考えられる。

5. AI 監査に関する今後の課題と提言

これまで AI サービスやシステムに関する監査を行うにあたって論点となる項目と、AI 監査が困難となる要因を整理してきた。これらをふまえ、本章ではこれらの課題を解決し、AI 監査を今後適切に進めていくための提言を行う。

5.1 AI 監査の制度設計の整備

現状国内においても安全、安心に AI を利用したいというニーズがある一方で、監査制度が未整備であることにより、有益な AI 監査の実施が浸透していない。国際社会に目を向けても、欧州 AI 規制法案の登場など、AI 監査に関する制度設計の必要性は今後高まってく

ることが想定されるため、日本においても AI 監査制度の設計に向けた議論が必要と考えている。AI サービスやシステムの監査に関する必要性は多くの政策文書などで触れられているものの、AI 監査の実現に向けては監査対象やタイミング等議論の必要な論点が多岐にわたる。そのため、制度設計の整備や、それに従った監査実施に向けて具体的な議論を行う際には本稿で整理したような共通理解をもとに議論を進めていく必要がある。

AI の外部監査の制度設計を検討する際には監査の品質と報酬のバランスの観点も考慮しなければならない。低すぎる監査水準、監査報酬は監査結果の品質そのものに疑義が生じる一方で、厳格な監査水準、監査報酬は AI サービス導入の参入障壁となり、イノベーションを阻害してしまう可能性がある。また、監査報酬という点においては、複数の組織が AI 監査の対象となった場合の監査費用の負担先についても考慮に入れた制度設計を検討する必要がある。

監査のタイミングについても AI の外部監査の制度設計においては考慮していく必要がある。従来のシステム監査においては、新規サービスや新規システムについてはリリース後のタイミングで監査を実施するケースが多い。一方で、被監査企業の立場から考えた場合、サービスやシステムのリリース前に問題点を解消し、万全の状態でリリースを迎えたいというニーズも考えられる。そのため、このようなリリース前の監査ニーズにも応えられる形での監査制度設計も検討していく必要がある。

さらに AI サービスやシステムは国や地域をまたいで開発、利用されることが多い。AI に対する規制が各国や地域ごとに整備されると、AI システムやサービスに対する監査を義務付けている国や地域と、そうではない国や地域の間で貿易障壁などが起きることも考えられる。このような観点からも AI 監査に関する議論を進めると同時にその実施基準の検討は相互運用が可能な形で行われることが望ましい。技術の進展が早いことも考慮に入れながら、サンドボックス制度[42]なども用いて AI 監査の制度設計を整備していくことが重要である。

5.2 AI 監査に関する人材の育成

AI 監査の実施には幅広いスキルや経験による多様な専門性が求められる。監査論の理解、被監査企業や提供しているサービスに関連する業界知識、AI に限らない IT 領域の知識や経験に加えて、AI 固有の技術的知見、倫理や文化、法規制等の非常に幅広い範囲での知識や経験が必要となる。一方で、これらの知識や経験を保有する人材はまだまだ国内においては少ないと想定され、今後の AI 監査のニーズに対応ができるような監査人の育成も進めていく必要がある。2 章にて解説した通り AI 監査の関係者と関係組織は多岐にわたることを考えると、人材育成の際には関連する知識や業務経験に加えて、各ステークホルダーと

[42]規制のサンドボックス制度とは「（略）新たな技術の実用化や新たなビジネスモデルの実施が、現行規制との関係で困難である場合に、新しい技術やビジネスモデルの社会実装に向け、事業者の申請に基づき、規制官庁の認定を受けた実証を行い、実証により得られた情報やデータを用いて規制の見直しに繋げていく制度」とされている（https://www.cas.go.jp/jp/seisaku/s-portal/regulatorysandbox.html）。

の適切な情報共有、複数人やチームによる監査体制のもとでのコミュニケーション力など、ソフトスキルも含めたより広範囲な観点での育成が求められる。また、国内外で監査実施者に対する様々な資格や認定制度があるが[43]、新たに AI 監査人としての個人、組織の資格要件制度が必要か否かについても検討が必要と考えている。

5.3 技術や利用の進展に伴う AI 監査のアップデート

AI に関連する技術研究は世界各国で盛んに行われており、日々新たなサービスやシステムが誕生している。また、AI サービスの利用形態や開発方法、関連者や関係組織についても、サービスやシステムの革新と連動してアップデートが行われている。このような変化に伴う新たなリスクや監査上の論点の登場、重要性の変化も想定されるため、陳腐化を避けるためにも AI 監査の制度や基準、手法についても併せてアップデートを行う必要がある[44]。また、アップデートの速度や頻度も重要な観点であり、国内外の相互運用性を鑑みながら、先進技術の進化のスピードに適したプロアクティブなアップデートを都度行えるよう、アップデートの仕組そのものについても今後併せての議論が求められる。

6. AI を安心して利用できる社会へ

本稿の目的は AI 監査に関する論点を整理し、関係者間で今後の AI 監査に関する議論を推進、発展させる共通の土台を構築することである。本稿では AI 監査において検討すべき論点はある程度整理を行ったので、今後は AI 監査のあるべき姿や実務上の取扱い等についての議論をさらに深めていく必要がある。

現在、AI 監査を取り巻く社会情勢として、2023 年 6 月 14 日に欧州議会で欧州 AI 規制法案の修正が採択[45]されたことに続き、2023 年 7 月 21 日には、米国ホワイトハウスが大手 AI 企業から、AI がもたらすリスクを管理するため、AI 技術を取り巻く広範な懸念に対処する主要な指針となる自発的なコミットメントを獲得したと発表した[46]。また、欧州評議会に設けられた「AI に関する委員会」では世界初の AI 条約を起草する交渉が行われている[47]。

---

[43] 例えば世界各国の公認会計士協会が定める公認会計士資格、経済産業省の定めるシステム監査技術者、内部監査人協会の定める公認内部監査人等が挙げられる

[44] 経済産業省の定める「システム監査基準」(https://www.meti.go.jp/policy/netsecurity/sys-kansa/sys-kansa-2023r.pdf)は 1985 年の策定後、1996 年、2004 年、2018 年、2023 年と改訂が繰り返されている。

[45] 欧州議会のサイトにて、欧州 AI 規制法案の修正が公開されている（https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence）。

[46] 米国ホワイトハウスのサイトにてコミットメントの詳細は公開されている（https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/）。

[47] 欧州評議会 AI 委員会（https://coe.int/en/web/artificial-intelligence/cai）のサイトにて、2023 年 7 月

このようにAIを提供する企業やAIを利用する企業でAIガバナンスやAIマネジメントが適切に行われているかを評価、監査によって把握する必要性がますます高まっている。AIの利用が今後より一層進展すると考えられる中で、本稿の論点整理がAIサービスやシステムを開発・提供する企業の責任あるAIの開発・運用を実現し、全ての人が安心してAIを利用できる社会の実現に寄与することを期待する。

また、AI監査をめぐる議論は、急速に発展し、多くの専門家の知見が不可欠である。またAI技術の発展も急速であるため、本政策提言もこのダイナミックに動く状況に応じて、最新の情報とベストプラクティスに基づいて充実化していくことが求められる。国内外の専門家や関連団体との協力により、より効果的で実用的なAI監査の議論の構築に資することを期待する。

7. 補章：AI監査と生成AI

本稿は冒頭に示したように、深層学習を含む機械学習をAI技術として扱い、その監査をめぐる論点と困難な点を整理してきた。一方、2022年以降急速に利用が拡大した生成AIを監査するにあたっては、本稿で挙げた論点や課題について共通のものが多いと考えているものの、生成AIに特有の論点や課題も想定されるため、本コラムで紹介する。なお、前提として本研究会において生成AIとは、基盤モデルや大規模言語モデル、その他の技術を活用してテキストデータや画像データ等の新しいデータを生成するAIであると捉えている。

7.1 生成AIの可能性と課題

本稿が対象としているAIはOECDの「人間が定義した一定の目的のために、現実または仮想の環境に影響を及ぼす予測、提言または判断を行うことができる機械ベースのシステム」と定義しており、ここに生成という単語は入っていない。

生成AIが認識や予測を行うAIと異なる点は、既存の画像、音声、動画、テキスト等のコンテンツやデータを学習して新しいコンテンツやデータを生み出すことである。しかし、生成AIは2022年以降に登場した全く新しい技術ではなく、従来の認識、予測、判断を行うAIと同様、機械学習手法の一種である。そして、その可能性と問題もすでに2010年代後半から議論されていた。例えば有名人の偽動画、過去の偉人の作風を模倣した作品の自動生成は行われており、問題点として指摘されていた。機械が生成することにより、時には人間では思いつかないような豊かな発想や表現も可能となる一方、著作権の問題、クリエイターや模倣の対象となった人の尊厳、フェイクニュースや名誉棄損等の法的、倫理的、社会的、経済的な課題は顕在化していた。

そのような状況と2022年以降の状況で何が異なるか。一つはAIサービスやシステムと

23日に統合版ドラフトが公開されている（https://rm.coe.int/cai-2023-18-consolidated-working-draft-framework-convention/1680abde66）。本会議の解説などについては東京大学未来ビジョン研究センターのイベント報告などを参照（https://ifi.u-tokyo.ac.jp/project-news/16287/）。

して無料・有料で広く一般利用者が利用しやすい形で提供されたことである。2022 年以前も一部の生成 AI サービスやシステムは利用可能であったが、一般利用者による利用は限定的であった。2022 年以降、一般利用者の利用拡大に伴い生成 AI の利用者は激増し、今後 AI 監査の潜在的対象に生成 AI を用いたサービスやシステムも増加する可能性があると考えられる。

7.2 生成 AI に重視される論点と課題

認識、予測等を行う AI と同様、生成 AI も機械学習の手法の一つであるため、本稿で扱われている AI 監査を巡る論点や AI 監査を困難にする要因は共有される。一方、生成 AI 固有あるいは生成 AI に特徴的な課題としてはいくつか考えられる。

例えば、2 章の AI 監査を巡る論点の項目においては、AI 監査の立証命題(2.2)に変化が生じる可能性がある。認識 AI や予測 AI では公平性、透明性等の観点を取り上げたが、生成 AI や対話 AI では著作権や偽情報や誤情報などのほか、個人の尊厳、感情操作、機械への依存など人と機械の関係性（ヒューマンマシンインタフェース）に関する議論も活発に行われている。AI を社会の価値や特定の領域に調整すること（アラインメント）自体はファインチューニングなどの手法で技術的には検証されているが、一方でこのような価値を適切に評価するための規準の開発も未熟であるため（3.3）、その点が生成 AI の監査を難しくさせると考えられる。

生成 AI の利用拡大に目を向けると、基盤モデル、大規模言語モデルやポータルなどのマシン性能も含めた生成 AI の利活用技術の急速な発展に加えて、対話型生成 AI というインタフェースの普及が 2022 年以降の文章や画像生成 AI を普及させた要因とも考えられている。そのため、AI 監査の対象(4.3)で議論されている内容に、インタフェースのデザインや設計、例えば機密情報を入力しないようにとする説明が適切に行われているか等が監査対象となりうる。同時にこのようなインタフェースは人と AI の相互作用が前提となっているために、AI 技術の複雑性(3.1)を増加させる。本稿で紹介した AI 監査においても、最終判断は人間が行うのか AI と人間の協働によって行われるのか等の前提条件が重要となることを指摘したが、対話型生成 AI の生成物の監査が必要になった場合は、どのような指示を出したのか等の記録や、入力された指示を制御するロジック等も監査対象となる可能性もある。例えば、不適切な入力指示を除外、補正して生成 AI からの出力をコントロールしている場合には、そのような除外や補正ロジックが適切に稼働していることも生成 AI を利用する上では重要な確認項目と考えられる。加えて、生成 AI の出力物に対する監査実施の際には、生成物に対して適切な表示が行われているかという点での監査も生成 AI に特徴的な監査手続として想定される。例えば、欧州 AI 規制法案においては生成 AI の出力コンテンツについては AI により生成されたことを表示することを求めているが、これらの要求に従っているかについて、実際に監査でウォーターマーキングに関する検証として確かめることも今後の論点の一つとして考えられる。

また、生成 AI に関してはコアとなるような大規模なモデルを個別の AI システムが参照

することで機能を実現しているケースが多い。例えば、文章生成の AI に関して言えば大規模言語モデルが相当する。さらに個別の AI システムの利用をサポートするような様々なツールも日夜開発されている状況である。このように、生成 AI のケースにおいては一つの AI サービスやシステムを利用するにあたって関係するシステム・ツールや関係者が多く、国や組織をまたぐケースも多発するため、関係者の全体把握がこれまで以上に困難となる可能性がある。

上記のような点から生成 AI に重点の置かれる課題もあるが、今後の技術発展に応じて監査の論点や手法に関しても継続的にアップデートしていくことが重要である。

参考文献

阿子島隆, 福田重遠. 2020a. AI システムの内部監査を考える(第 1 回)AI 活用の動向と課題. 旬刊経理情報, *1592*, 50–54.

阿子島隆, 福田重遠. 2020b. AI システムの内部監査を考える(第 2 回)AI システム監査の全体像. 旬刊経理情報, *1595*, 56–60.

阿子島隆, 福田重遠. 2020c. AI システムの内部監査を考える(第 3 回・完)AI システム監査の実施モデル. 旬刊経理情報, *1596*, 46–52.

鳥羽至英. 2009.『財務諸表監査　理論と制度【基礎編】』国元書房.

山浦久司. 2008.『会計監査論』中央経済社.

独立行政法人情報処理推進機構.2013.『共通フレーム 2013』独立行政法人情報処理推進機構.

秋元一郎.2022.『AI ビジネス大全』プレシデント社.

Bandy, J. 2021. “Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits”, Proceedings of the ACM on Human-Computer Interaction, Vol.5, No.CSCW1,Article 74.pp.1-34.

de Boer, M. 2023. “Trustworthy AI and accountability: yes, but how? What the EU AI Act’s approach to AI accountability can learn from the science of algorithm audit”. [Thesis, externally prepared, Universiteit van Amsterdam].

Batarseh, F. A., Freeman, L., and Huang, C.-H. 2021. “A survey on artificial intelligence assurance”, Journal of Big Data, Vol.8, 60.

Dwork, C. 2012. “Fairness Through Awareness”, Proc. of the 3rd Innovations in Theoretical Computer Science Conf: 214-226.

Hardt, M. 2016. “Equality of Opportunity in Supervised Learning”, Advances in Neural Information Processing Systems 29: 3323-3331.

Institution of Internal Auditors (IIA), 2017a. “Artificial Intelligence – Considerations for the Profession of Internal Auditing,” Global Perspective and Insights (Special Edition).

IIA, 2017b. “The IIA’s Artificial Intelligence Auditing Framework Practical Applications, Part A,” Global Perspective and Insights (Special Edition),
IIA, 2017c. “The IIA’s Artificial Intelligence Auditing Framework Practical Applications, Part B,” Global Perspective and Insights (Special Edition).
IIA, 2017d. “International Standards for the Professional Practice of Internal Auditing (IPPF)”（内部監査人協会「内部監査の専門職的実施の国際基準」『専門職的実施の国際フレームワーク-2017 年版-』）.
IIA, 2020. “The IIA’s Three Lines Model: An update of the Three Lines of Defense”(IIA Position Paper).
Issa, H., T. Sun, M. Vasarhelyi, 2016. “Research Ideas for Artificial Intelligence in Auditing: The Formalization of Audit and Workforce Supplementation,” Journal of Emerging Technologies in Accounting, Vol. 13, No. 2, pp.1-20.
Jobin,A., Ienca, M., Vayena, E. 2019. “The global landscape of AI ethics guidelines,” Nature, Machine Intelligence, 1, 389-99.
Koshiyama, A., E. Kazim and P. Treleaven, 2022. “Algorithm Auditing: Managing the Legal, Ethical, and Technological Risks of Artificial Intelligence, Machine Learning, and Associated Algorithms”, Computer, Vol.55, No.4, pp.40-50.
Mökander, J., J. Schuett, H. Kirk, L. Floridi, 2023, “Auditing Large Language Models: A Three-Layered Approach,” arXiv:2302.08500.
Nakano, M., 2023. “Toward Building the Framework for Artificial Intelligence Audit Theory”, Industrial Technology No.45, pp.52-55.
New York City Local Law 144, 2021.

謝辞

本研究は東京大学未来ビジョン研究センターフラッグシッププロジェクト「AI 社会における未来ビジョンのデザイン」のほか「東京大学卓越研究員」制度による研究の一環として実施した。また、AI ガバナンスプロジェクトにおける企業や行政の方々との共同研究の成果も本提言の一部に含まれる。

また、本政策提言を執筆するにあたり、多くの方に有益な助言をいただいた。時間や所属組織の関係上、お名前の記載が叶わなかったが、有益なフィードバックを寄せていただいた方々がいることも付け加え、フィードバックを寄せていただいた以下の方々に感謝申し上げる。

阿子島隆（システム監査学会会員　公認内部監査人（CIA））
市原直通（EY 新日本有限責任監査法人 AI リーダー）
伊藤公一（PwC あらた有限責任監査法人アシュアランス・イノベーション＆テクノロジー部 パートナー/ AI 監査研究所 副所長）

宇宿哲平（有限責任あずさ監査法人 Digital Innovation 事業部 パートナー）
加藤信彦（EY 新日本有限責任監査法人 アシュアランスイノベーション本部イノベーション戦略部／AI ラボ部長）
島田裕次（東洋大学工業技術研究所 客員研究員）
清水希理子（PwC あらた有限責任監査法人アシュアランス・イノベーション＆テクノロジー部(AIT) シニアマネージャー）
城山英明（東京大学未来ビジョン研究センター 教授）
瀧博（立命館大学経営学部 教授）
仲浩史（東京大学未来ビジョン研究センター 教授/内部監査人協会(IIA)グローバルボード理事）