

グローバルな視点と洞察 特別版-A | 第3部

# IIAの人工知能監査フレームワーク 実務への適用 パートB

内部監査人協会（IIA）

訳者：堺 咲子

内部監査財団（Internal Audit Foundation） 評議員  
インフィニティコンサルティング 代表  
CIA, CCSA, CFSA, CRMA, CPA (USA)

## 目次

IIAのAI監査『フレームワーク』…… 20	ブラックボックス…… 22
人的要因…… 20	ガバナンス…… 22
倫理…… 20	成果測定…… 22
アルゴリズムバイアス…… 20	最後に…… 23
意味理解…… 21	

### 諮問委員会

#### IIA マレーシア

CIA, CCSA, CFSA, CGAP, CRMA

ヌル・ハヤティ・バハルディン氏

#### IIA アフリカ地域連合

CIA, QIAL

レセディ・レセテディ氏

#### IIA オランダ

CIA, CCSA, CGAP

ハンス・ニューランド氏

#### IIA アラブ首長国連邦

CIA, CCSA, CRMA

カレム・オベイド氏

#### IIA 北米

CIA, CRMA, CPA

キャロライン・セイント氏

#### IIA コロンビア

CIA, CCSA, CRMA

アナ・クリスティーナ・ザンブラノ・プレシアド氏

Copyright © 2017 by The Institute of Internal Auditors, Inc., (“The IIA”) strictly reserved. Any reproduction of The IIA name or logo will carry the U.S. federal trademark registration symbol ®. No parts of this material may be reproduced in any form without the written permission of The IIA.

Permission has been obtained from the copyright holder, The Institute of Internal Auditors, 1035 Greenwood Blvd., Suite 401 Lake Mary, FL 32746, U.S.A., to publish this translation. No part of this document may be reproduced, stored in any retrieval system, or transmitted in any form, or by any means electronic, mechanical, photocopying, recording, or otherwise, without prior written permission of The IIA.

## 注

本文書は3部シリーズの第3部である。詳細は「人工知能－内部監査の専門家が考慮すべきこと」（編集注：『月刊監査研究』2018年2月号掲載）および「IIAの人工知能監査フレームワーク－実務への適用パートA」（編集注：『月刊監査研究』2018年3月号掲載）を参照。

## はじめに

想像して欲しい。倫理委員会や調査委員会の面前に、ロボットが、または少なくともその頭脳がいるところを。それは起こり得るのである。ザ・ガーディアン紙<sup>1</sup>によると、ロボットが公共空間に入って人間と一緒に働き始めているので安全対策の必要性が一層高まってきている、と学者たちは主張している。科学者たちは、ロボットが自らの意思決定を追跡するための「倫理的ブラックボックス」を装備して、事故が起こった時に行動を「説明」できるようにしようと試みている。

オックスフォード大学の教授たちは、航空機墜落事故調査にブラックボックスやコックピットのボイスレコーダーを持ち込んで、悲劇から得た安全上の重大な教訓としている航空業界の例と規則にロボット工学企業も倣うべきだ、と主張している。ロボットに装備された倫理的ブラックボックスは、ロボットの意思決定、その意思決定の根拠、ロボットの動き、さらにカメラ、マイク、距離計などの

センサーからの情報を記録するだろう。

これらの動きは、ナイトスコープ<sup>2</sup>社のK5警備ロボット「スティーブ」が転倒して勤務中に噴水に落ちた時<sup>3</sup>や、別のK5ロボットがカリフォルニア州マウンテンビューの通りをパトロール中に41歳の男性に暴行された時<sup>4</sup>や、スタンフォードのショッピングセンターで警備ロボットが暴走して1歳4か月の幼児を轢いたと非難された時のような、最近の事件の結果を受けて起こったものである。なお、関係者の怪我は全て軽傷だった。

人工知能の倫理については何百ものレポートがあるが、大部分は軽い内容であり、人間を第一に考えるという面白みのないものばかりである、と「バックチャネル」編集者のスコット・ローゼンバーグ氏は書いている。ローゼンバーグ氏は、ニューヨーク大学のAI Now研究所が最近公表した「信頼できる公平な結果を保証することなくAIで社会を再構築しよう」と試みているハイテク産業」に関するレポート<sup>5</sup>から引用している。1つのレポート<sup>6</sup>は、「AIを倫理基準で縛ろうとした今までの努力は大失敗しているので、AIの新しい倫理的枠組みは、強力な産業的、行政的、軍事的利害を持つ者がAIを設計して採用する際に責任を負うように、個人の責任を超えたものにする必要がある」と結論づけている。著者の意見によると、AIシステムは警備活動、教育、ヘルスケア、その他多様で脆弱な分野で導入されており、「アルゴリズムの失敗が人命を脅かすこともあり得る」。

<sup>1</sup> 訳者注：<https://www.theguardian.com/science/2017/jul/19/give-robots-an-ethical-black-box-to-track-and-explain-decisions-say-scientists>

<sup>2</sup> 訳者注：<https://www.knightsscope.com/about/>

<sup>3</sup> 訳者注：<https://www.theguardian.com/technology/2017/jul/18/robot-cop-found-face-down-office-block-fountain>

<sup>4</sup> 訳者注：<http://abc7news.com/technology/police-drunk-man-knocked-down-300-pound-robot-in-mountain-view/1915713/>

<sup>5</sup> 訳者注：[https://assets.contentful.com/8wprhvhnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/\\_AI\\_Now\\_Institute\\_2017\\_Report\\_.pdf](https://assets.contentful.com/8wprhvhnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/_AI_Now_Institute_2017_Report_.pdf)

<sup>6</sup> 訳者注：<https://www.wired.com/story/ai-experts-want-to-end-black-box-algorithms-in-government/>

## IIAのAI監査『フレームワーク』



「人工知能－内部監査の専門家が考慮すべきこと」で説明した通り、AIにおける内部監査の役割は「人工知能が短期、中期、または長期の価値を創造する組織の能力に否定的または肯定的な影響を与える程度を評価し、理解し、伝達するのに役立つこと」である。

内部監査がこの役割を果たす手助けとして、内部監査人は組織に合わせてAI関連のアドバイザー、アシュアランス、またはアドバイザーとアシュアランスが融合したサービスを提供する際にIIAのAI監査『フレームワーク』を活用することができる。この『フレームワーク』は、AI戦略、ガバナンス、人的要因の3つの包括的な要素と、7つの構成要素であるサイバーレジリエンス、AI能力、データ品質、データアーキテクチャとインフラストラクチャ、成果測定、倫理、ブラックボックスで構成されている。

内部監査は『フレームワーク』を使って組織のAI活動に関連するアドバイザー、アシュアランス、またはアドバイザーとアシュアランスが融合した内部監査サービスを提供する際に、多くの監査目的またはコントロール目的、および活動または手続を考慮すべきである。『フレームワーク』の戦略（サイバーレジリエンス、AI能力）とガバナンス（データアーキテクチャとインフラストラクチャ、データ品質）に関連する目的、および活動または手続については、「IIAの人工知

能監査フレームワーク実務への適用 パートA」で説明した。本文書では、人的要因（倫理、ブラックボックス）とガバナンス（成果測定）に関連する目的、および活動または手続を説明する。

### 人的要因

倫理とブラックボックスを含む人的要因は、期待した結果を提供しようとするAIの能力に支障を来す人的ミスを取り上げている。

### 倫理

人的ミスや偏った判断（ともに意図的であってもなくても）をする人間によって開発されたアルゴリズムは、そのアルゴリズムの性能に影響する。人的要因として以下を検討すること。

- 意図的でない人間の判断の偏りがAI設計に影響する、というリスクが特定され管理されているか。
- AIの結果が元の目的を反映していることを確かめるために効果的にテストされているか。
- 技術的な複雑さを考慮すると、AIには透明性があるか。
- AIのアウトプットは合法的に、倫理的に、責任ある形で使用されているか。

### アルゴリズムバイアス

マッキンゼー社の最近のレポートによると、企業はビジネス上の意思決定に機械学習を素早く適用している。そのプログラムは、頻繁に更新される大規模なデータセットを扱うための複雑なアルゴリズムを設定する。しかし、アルゴリズムバイアスが見落とされてチェックされないままになっていると、機械学習の目的に影響を及ぼす可能性があり危険である（マッキンゼー社の「Controlling machine-

learning algorithms and their biases<sup>7</sup>」を参照）。

例えば信用格付では、延滞や債務不履行なく長期間融資を維持している顧客は通常「低リスク」と判断される。しかし、この顧客の住宅ローンが期限付きの相当な優遇税制に支えられてきたことは知られていない可能性がある。やがて債務不履行が急増しそうなことは、貸付機関の統計的リスクモデルでは説明されない可能性がある。適切なデータにアクセスして専門家の助言を受ければ、機械学習の予測モデルはデータの中の隠れたパターンを見つけてそのような急増をなくせるかもしれない。さらにビジネス上の意思決定以外では、アルゴリズムバイアスが市民に問題や不安を与えるような間違いを起こす可能性がある。例えば、グーグル社のフォトサービスや類似のサービスは、人間、物体、風景を識別するために使用されるが、人種に関する配慮が欠けていた場合や、将来の犯罪者を予測するためのリスクアセスメントに使用されたソフトウェアが偏った判断を示した場合は、非常に恐ろしい状況になり得る。

### 意味理解

マシンにできることには限界があるので、人間はAIのアウトプットを理解できなければ

ならない。マッキンゼー社のレポートで述べられているように、「AI時代における意味理解は、マシンにできることとできないことを理解することから始まる。例えばマシンは、ある種の診断を人間よりも正確に行える可能性がある。しかし、患者がその意味を理解して影響に対処するのを支えるのは、看護師や医師やセラピストの責任である。それこそが知識と意味理解との違いである」。このレポートによると、コーディング、分析、データサイエンスなどのハードスキルはAIにとって重要だが、協力、共感、意味理解などのソフトスキルも重要である（「McKinsey Quarterly 2017 Number 4」56-61頁参照）。

IIAが示した関連する目的、および活動または手続は、監査計画を規定するものではないが、示した例は監査目的またはコントロール目的の特定、およびAI監査の計画と実施に役立つはずである。

AI監査は、IIAの「基準2200：内部監査（アシュアランスおよびコンサルティング）の個々の業務に対する計画の策定」に適合すべきである。AI監査計画とAI監査目的および手続は、組織のニーズに合わせて常にカスタマイズすべきである。

倫理に関連する目的、および活動または手続	
監査目的またはコントロール目的	活動または手続
組織のAI活動の成果に意図しないバイアスがないというアシュアランスを提供する。	AI活動の意図した結果（戦略目標）をレビューして実際の結果と比較する。差異が検出された場合はバイアスが原因であるかを判断する。
組織はAIのアウトプットの「意味理解」ができる。	AIのアウトプットとそのアウトプットから得られた意味をレビューする。

<sup>7</sup> 訳者注：<https://www.mckinsey.com/business-functions/risk/our-insights/controlling-machine-learning-algorithms-and-their-biases>

<sup>8</sup> 訳者注：<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

## ブラックボックス

メリアム・ウェブスター社のオンライン辞書によると、ブラックボックスとは「大抵は複雑な電子装置であり、その内部メカニズムは通常ユーザーから隠されているかユーザーには不可解である。広義には、不可解で未知の内部機能やメカニズムを持つもの」である。組織がタイプⅢやタイプⅣのAI技術（自習や相互通信できるマシンやプラットフォームを利用する技術）に移行するにつれて、アルゴリズムがどのように動いているかが分かりにくくなったり理解できなくなったりする。組織のAI活動がより高度化するにつれて、ブラックボックスはより大きな課題になる。

## ガバナンス

AIガバナンスとは、組織のAI活動を指示、管理、監視するために導入された構造、プロセス、手続を指す。組織のAIガバナンスの形式と構造は、当該組織の個々の特性によって異なる。AIガバナンスは、

- アカウンタビリティ、責任、監督体制を確立する。
- AI責任者が必要なスキルと専門知識を確保するのを支援する。
- AI活動やAIに関連する意思決定と活動が、組織の価値観、および倫理的、社会的、

法的責任に沿ったものとなるように支援する。

本シリーズの第2部「IIAの人工知能監査フレームワーク実務への適用 パートA」では、ガバナンスに関連してアカウンタビリティ、責任、監督、規制の遵守、データアーキテクチャとインフラストラクチャ、データ品質を取り上げた。本文書では成果測定を取り上げる。

## 成果測定

内部監査は、組織のAI施策の成果を測定するために不可欠な位置づけにある。計画段階で内部監査は、経営陣と取締役会にとって十分に信頼性の高い関連性のある有用な情報を提供するような指標を定める方法について助言することができる。ただし、内部監査はAI成果指標の設定や所有の責任を負ってはいならない。AIが導入されている組織の内部監査は、AIに関する第1のディフェンスラインのコントロールと第2のディフェンスラインの監督についてアシュアランスを提供すべきである。内部監査のAI能力を実証するには、AI監査にロボット工学によるプロセス自動化などのAI技術を使用すること以上に良い方法はないだろう。

### ブラックボックスに関連する目的、および活動または手続

監査目的またはコントロール目的	活動または手続
「ブラックボックス」データ（例：根底にあるアルゴリズム、内部機能、AIを可能にするメカニズム）に対する組織の理解度を評価する。	AIの開発と導入に関する方針、プロセス、手続をレビューして、ブラックボックスデータが特定されているかを検証する。 AIの成果の責任者にインタビューして、ブラックボックスデータを理解しており説明できるということを検証する。

成果測定に関連する目的、および活動または手続	
監査目的またはコントロール目的	活動または手続
A I 指標の設定方法について助言する。	A I 指標の設定責任者のための作業部会のファシリテーションを行う。十分に信頼性の高い関連性のある有用な情報という用語の重要性と意味を伝える。
A I の脆弱性に関するストレステストを行う。	極端な状況で A I がどのように機能するかを判断するために、銀行業界で使用されるストレステスト技法を適用する。
A I 関連監査の結果を伝達する。	A I 関連監査の結果を以下に適合するように伝達する。 <ul style="list-style-type: none"> <li>■ I I A 基準2400：結果の伝達</li> <li>■ I I A 基準2410：伝達の規準</li> <li>■ I I A 基準2420：伝達の品質</li> <li>■ I I A 基準2421：誤謬および脱漏</li> <li>■ I I A 基準2430：「『内部監査の専門職的实施の国際基準』に適合して実施された」旨の表現の使用</li> <li>■ I I A 基準2431：「基準」等に不適合な場合の内部監査（アシュアランスおよびコンサルティング）の個々の業務の開示</li> <li>■ I I A 基準2440：内部監査の結果の周知</li> </ul>
A I に関する第1のディフェンスラインのコントロールと第2のディフェンスラインの監督についてアシュアランスを提供する。	A I 関連監査にロボット工学や他の A I を積極的に活用する。

## 最後に

内部監査の専門家は、次のデジタル領域フロンティアになり得る人工知能に後れを取ることにはできない。内部監査人はこれに備えて A I の基本、内部監査が果たせる役割と果たすべき役割、および A I のリスクと機会を理解しなければならない。

組織の A I 技術と活動が、進んだ技術を使って社内で開発されたものであろうと第三者

によって開発された技術であろうと、内部監査は第1と第2のディフェンスラインと連携して取締役会と上級経営者に助言する準備をすべきであり、さらに A I のリスクマネジメント、ガバナンス、コントロールに対するアシュアランスを提供すべきである。内部監査人はこれらの課題に対処するために、IIA の人工知能監査の『フレームワーク』を活用すべきである。