

AI 監査と不正の多様性：監査人の責任に係る
経済実験

田口 聡志

**The effects of AI use and the diversity of frauds on jurors'
evaluations of auditor Legal liability: An experimental study**

Satoshi Taguchi

ITEC Working Paper
21-01
February 2021

AI 監査と不正の多様性：監査人の責任に係る経済実験

The effects of AI use and the diversity of frauds on jurors' evaluations of auditor Legal liability:

An experimental study

田口聡志 (同志社大学商学部)

Abstract : 本研究は、監査人が監査業務に AI を利用すること、および不正の多様性（不正の原因の違い）が、会計不正を見逃した場合の監査人の法的責任にどのような影響を与えるかについて、場面想定法を用いた陪審員の意思決定実験により明らかにする。実験は、アーカイバル・データがまだない未来の分析をする上で有効なツールである。特に本研究では、監査に対する社会的受容(social acceptance)を検証する手法として近年の監査研究で注目されている「陪審員実験」（被験者に陪審員役を課し、想定シナリオのもと監査人の法的責任を判定させる実験）により、日本公認会計士協会近畿会 IT 委員会の協力により、日本在住の公認会計士 188 名を被験者に、「AI 利用の有無」（利用あり vs. 利用なし）と「不正の原因」（古典的不正 vs. 会計上の見積り判断による不正）を考慮した 2×2 between-participant design のサーベイ実験をおこなった。

その結果、①AI の利用は、監査人の責任をより重くする恐れがあること、また、②特に AI を利用する場合、会計上の見積り判断による不正を見逃した際に、監査人の責任はより重くなる可能性があることが明らかにされた。前者①は、AI を利用した監査の逆効果といえる。また後者②は、AI 監査において、監査人は今まで以上に見積り判断など人間が得意とする（AI が不得意とする）項目に注力した監査を進める必要があることを示唆するものである。特に現在は、COVID-19 問題を背景として企業の DX(Digital Transformation)化が進んでいるが、本稿はテクノロジーと経済社会の共存のあり方の議論に対しても一定の示唆を与える。

Keywords: AI 監査; 社会的受容; 不正の多様性; 陪審員; 会計上の見積り

(注) 本研究の内容は、今後英語化し英文ジャーナル投稿を予定しており、本稿は、その準備のためのディスカッション用に preliminary report として纏めたものである。

I. INTRODUCTION

現在、AI(Artificial Intelligence)など新しいテクノロジーが人の業務に与える影響が議論されており(Autor and Dorn 2013; Frey and Osborne 2017; McAfee and Brynjolfsson 2018)、すでに医療や法律の分野においてAIの利用がはじまっている(Murdoch and Detsky 2013; Rissland, Ashley and Loui 2003)。監査業界においても、相次ぐ会計不祥事を受け、大手監査法人が、AIを利用した監査(AI監査)に取り組みはじめている(日本経済新聞 2018年10月8日付朝刊)。たとえば、EY新日本有限責任監査法人は、AIによる会計仕訳の異常検知技術の特許を取得するなど、先んじてAI監査に取り組んでいるようである¹。この背景には、新しいテクノロジーの進展だけでなく、大型会計不正により、監査の社会的意義が問われていること(監査の品質向上の社会的要求)、および原則主義を採用する国際会計基準(IFRS: International Financial Reporting Standards)の登場により、監査人の実質判断が問われていることが挙げられる。

ここで、もし仮にAIを利用した監査において、監査人が不正を見逃した場合、監査人の責任は軽減されるのだろうか。また、不正の原因の違いは、AI監査における監査人の責任に影響を与えるだろうか。制度面においては、PCAOBを中心に、監査人の他の専門家の利用に関するルールが議論されている。たとえば、AS(Audit Standard) 1210: Using the Work of a Specialistでは、海外子会社の連結(別のaudit firmが監査を担当する場合)や資産負債の公正価値評価を背景に、他の専門家を利用した場合の監査人の責任についての議論がなされている。また、原則主義であるIFRS適用により会計上の見積り要素が増加し、「人の要素」が占める割合が大きくなると、経営者の見積り・判断と監査人のそれとが食い違う可能性も広がる可能性がある。その傾向は、昨今の監査報告書の改定(Key Audit Matters導入)の中で、より強くなるものと考えられるし、AI監査においても、見積り部分については、人の監査人が担うウェイトは未だ大きい(AIが会計上の見積りを検証することには限界があり、AIでは代替し得ない領域)といえるかもしれない。

先行研究は、監査人がdecision aidを利用した場合の監査人の責任を、陪審員役の被験者に判定させる実験がなされている(Lowe, Reckers and Whitecotton 2002; Grenier, Lowe, Reffett and Warne 2015)。また、先行研究の一部は、これをAI監査に係る監査人の責任問題に拡張している(Taguchi 2018)。本研究は、さらにこれらを拡張し、会計不正の原因の違い(「タイプの多様性」)に着目する。これまで、会計基準の質の違い(原則主義 vs. 細則主義)が監査人の責任に与える影響を取り扱う実験研究はあるが(e.g. Grenier, Pomeroy and Stern 2015;

¹ EY新日本有限責任監査法人のプレスリリース website (<https://www.shinnihon.or.jp/about-us/news-releases/2019/2019-01-18.html>) を参照(2021年2月1日閲覧)。

Kadous and Mercer 2012; Kadous and Mercer 2016)、不正のタイプの多様性(原因の違い)に着目したり、またそれを AI 監査と組み合わせた研究は皆無である。

そこで本研究は、監査人が AI を利用することが、監査人の不正を見逃した場合の責任にどのように影響を与えるか、また不正の多様性(不正の原因の違い)がどのような影響を与えるか、場面想定法を用いた陪審員の意思決定実験により明らかにする。実験は、アーカイバル・データがまだない未来社会の分析をする上で有効なツールである。特に本研究では、監査に対する社会的受容(social acceptance)を検証する観点から、近年の監査研究で注目されている「陪審員実験」(被験者に陪審員役を課し、想定シナリオのもと監査人の責任を判定させる実験)により、日本の公認会計士 188 名を被験者に、「AI 利用の有無」(利用あり、利用なし)と「不正の原因」(古典的な不正の場合と、会計上の見積り判断による不正の場合)とを考慮した 2×2 between-participant design の実験をおこなった。

その結果、①AI の利用は、監査人の責任をむしろ重くする可能性があること、また、②特に AI を利用する場合、会計上の見積り判断による不正を見逃した際に、監査人の責任はより重くなる可能性があることが明らかにされた。前者は、AI 監査の逆効果といえる。また後者は、AI 監査において、監査人は今まで以上に見積り判断など人間が得意とする(AI が不得意とする)項目に注力した監査を進める必要があることを示唆する。

本研究は、AI 時代の監査のあり方について実務的な示唆を与える。本研究の結果から、社会の目は、AI 監査における監査人の責任をより厳しくみていることが理解できる。このことは逆に、AI 監査において、監査法人が社会に対して積極的に情報発信することの重要性を示唆している。たとえば、現状では、AI 監査によって、監査の質が向上することや、不正の量が減少することは、あまり社会に向けてアピールされていない。これらを社会に発信するだけでなく、制度的にも監査における AI 利用のガイドラインを整備・明確化し、監査において、どのような点で AI が利用され、またどのような点で人間の監査人が判断をおこなうのかを、社会に伝える制度的努力が求められる。

また、本研究は、AI の社会的受容(social acceptance)の理論にも一定の示唆を与える。AI の社会的受容については、実験社会科学(ESS: Experimental social science)や科学技術社会論(STS: Science and Technology Studies)などで様々な議論がなされている。たとえば、自動運転の倫理ジレンマ研究の成果(Awad et al. 2018; Bonnefon et al. 2016)によれば、人間は、AI の意思決定が(人間の)既存の社会ルールに沿ったかたちでなされるべきであるという無意識のバイアスを有していることが明らかにされている。これに対して、本研究は、AI を人間が利用する場合の人間のあり方に焦点をおく点で、これまでの先行研究とは違う視点から社会的受容の問題を捉えている。また、本研究は、AI を利用することで、人間のなすべき役割

の重要性が減じるのではなく、むしろ高まるということを示しており、この点は、AI と人間とが共存する社会を構築するうえでの重要な示唆といえる。特に現在は、COVID-19 問題を背景として企業の DX(Digital Transformation)化が進んでいるが、テクノロジーと経済社会の共存のあり方の議論に対しても、本稿は一定の示唆を与える。

本稿の構成は、以下のとおりである。II で先行研究と仮説を示し、III で実験計画について述べ、IV ではその結果を示す。最後にVで、結果に対する議論をおこなう。

II. RELATED LITERATURE AND HYPOTHESES

制度面において、現在、PCAOB を中心に、監査人の他の専門家の利用に関するルールが議論されている。たとえば、AS (Audit Standard) 1210: Using the Work of a Specialist では、海外子会社の連結（別の audit firm が監査を担当する場合）や資産負債の公正価値評価を背景に、他の専門家を利用した場合の監査人の責任についての議論がなされている。

研究面においても、近年、監査に対する「社会の目」や社会的受容 (social acceptance) を測る手法として、被験者を陪審員 (Juror) 役として、様々なケースをもとに監査人の法的責任を判定させる心理実験が注目を浴びている (e.g. Grenier, Reffett, Simon, and Warne 2018)²。監査人の Decision-Aid 利用時の法的責任について、Lowe et al. (2002)は、「Decision aid 自体の信頼性の高低」（2 レヴェル：信頼性が高い vs. 信頼性が低い）と「監査人による Decision aid の利用状況」（4 レヴェル：no decision aid vs. override vs. relied on incorrect decision aid vs. available but not used）という要因が、陪審員による監査人の法的責任の判定にどのような影響を及ぼすかを実験により検証している。その結果、監査人の利用状況が法的責任に影響を与えるのは、decision aid の信頼性が高い場合に限定され、かつ過剰利用した場合にのみ、責任が重くなるということが明らかにされている³。

AI 監査における監査人の責任について検証した研究としては、Taguchi (2018)がある。Taguchi (2018)は、陪審員実験を用いて、監査の失敗に対する責任を、「人間の監査人条件」、「人の監査人が AI 利用条件」、「Android 監査人条件」という3つの条件間で比較している。

² たとえば、米国における監査人の民事責任について、重要な事件は陪審員制で決していることから、近年の米国を中心とする監査研究では、被験者が陪審員役として、仮想的な監査の訴訟シナリオにおける陪審員としての判断や意思決定をおこなう実験研究が増えてきている。これは、監査の社会的受容、つまり「社会の目」が監査をどのように考えているか、監査の問題をどのように評価するかという点と繋がるものである。

³ また、会計基準の質の違い（原則主義 vs. 細則主義）が監査人の責任に与える影響を取り扱う実験研究としては、Grenier, Pomeroy and Stern (2015)、Kadous and Mercer (2012)、Kadous and Mercer (2016)などがある。

そして、AI利用の監査の場合の監査人の責任のほうが、人間のみの監査の場合の監査人の責任よりも重くなるという結果を示している。この点について、さらに Taguchi (2018) は、被験者の意思決定理由を分析しており、その結果、AI 利用監査における「AI 選択責任」（AI を利用する意思決定をしたこと自体に対する責任）が付加されることによるものであることが明らかになっている。本研究は、Taguchi(2018)を、不正の原因の違いを加味したセッティングに拡張する。

ここで、原則主義である IFRS 適用により会計上の見積り要素が増加し、「人の要素」が占める割合が大きくなると、経営者の見積り・判断と監査人のそれとが食い違う可能性も広がる可能性が高い。そうすると、監査人の責任の判定は、ますます複雑化するものと考えられる。また、その傾向は、昨今の監査報告書の改定（Key Audit Matters 導入）の中で、より強くなるものと考えられるし、AI 監査においても、たとえば減損損失の計上など見積り部分については、人の監査人が担うウェイトは大きい（AI には代替し得ない部分）といえるかもしれない。監査人の責任の判定はさらに複雑化する可能性が高い。これらの点について、たとえば、先行研究では、会計基準の質の違い（原則主義 vs. 細則主義）が監査人の責任に与える影響（e.g. Grenier, Pomeroy and Stern 2015; Kadous and Mercer 2012; Kadous and Mercer 2016）、KAM ないし CAM(Critical Audit Matters)が監査人の責任に与える影響（Brasel, Doxey, Grenier, and Reffett 2016; Gimbar, Hansen, and Ozlanski 2016）を分析した実験研究はあるが、不正の原因の違いを AI 監査のセッティングに拡張した研究は皆無である。

ここで、不正の原因が、減損損失の計上など人間の判断や見積りに大きく依拠するものである場合には、たとえ AI 監査であったとしても、人間の監査人が対処すべき点であるという考えから、これを見逃した監査人の責任はより重く判定されるものと予想される。他方、不正の原因が、いわゆる古典的手法によるもの（たとえば、売上の過大計上など通常のルーチン業務に係るもの）であれば、AI 監査においては、AI が対処すべき点であるという考えから、これを見逃した監査人の責任はより重く判定されることはない予想される⁴。

以上から、本研究では、以下の2つの仮説において、それらを実験で検証する。

⁴ なお、これらの背後には、AI 監査における「業務のシェアと責任のシェア」（業務をシェアするなら、責任もシェアすべき）という発想が暗黙のうちに存在するといえる。ただし、業務と責任の関係をどのように捉えるかについては、実は社会的厚生最大化の観点から慎重に考える必要がある。これについては、田口(2018a)を参照。

仮説1. AI 利用に係る監査人の責任 (人のみの監査(control) vs. AI 監査):

「監査の失敗」に関する監査人の責任は、人のみの監査の場合よりも、AI 監査の場合のほうがより重くなる。

仮説2. 不正原因の違いと AI 監査の責任 (見積り不正 vs. 古典的不正):

不正の原因の違いは、AI 監査人の責任に以下のような影響を与える。

- 古典的手法による不正→AI が対処すべき点→監査人の責任↓
- 判断・見積りによる不正→人間が対処すべき点→監査人の責任↑

III. METHOD

構造: 本研究では、被験者に陪審員役を課し、想定シナリオのもと「監査の失敗」時の監査人の法的責任を判定させる実験により、「AI 利用の有無」(利用あり vs. 利用なし)と「不正の原因」(古典的不正(架空売上) vs. 会計上の見積り判断による不正(減損損失の計上))とを考慮した 2×2 between-participant design の実験をおこなっている。図表 1 は実験の構造である。

図表 1 実験の構造

		監査人の AI 利用	
		なし	あり
不正の原因	見積り・判断 (減損損失計上)	人のみ監査 見積り不正	AI 監査 見積り不正
	古典的手法 (架空売上)	人のみ監査 古典的不正	AI 監査 古典的不正

被験者の意思決定: 実験の被験者は、図表 1 に示される 4 つの条件のうち 1 つだけに参加し、Juror としての役割を担う。具体的には、場面想定法により、被験者は、仮想シナリオを読んだ後に、監査の失敗に対する監査人の責任の重さやその理由をリカートスケールで回答する(図表 2) ⁵。

⁵ 会計の心理実験において「意思決定の強さ」をリカート・スケールで回答することの合理性は、Jackson (2008)、Jackson, Rodgers and Tuttle (2010)、Seybert (2010)などを参照。

図表 2 被験者の意思決定変数

- | |
|-----------------------------------------------|
| (i) 監査人の責任 (0-100 ポイント、リカートスケール) |
| (ii) 意思決定の理由を示す各質問に関する同意度 (1-7 ポイント、リカートスケール) |

実験シナリオ：本研究では、図表 1 の実験構造をもとに、全 4 パターンの仮想シナリオを用意している。シナリオ作成にあたっては、日本公認会計士協会近畿会 IT 部会の協力を得て、現実の監査人からみても理解不能な点がないように配慮している。

全 4 パターン共通のシナリオとして、監査対象企業が倒産したこと、倒産後に、企業の不正が明らかとなり、監査人はそれを見抜けなかったこと（「監査の失敗」が起こってしまったこと）が示される。そして最後に、その不正の手法が示される。

AI 利用の有無は、シナリオの冒頭で操作される。AI 利用あり条件（「AI 監査」条件）では、シナリオの最初に、監査に AI が導入されている旨、および、その AI 監査のイメージが付加される（他方、AI 利用なし条件では、このような記述はない）。

不正の多様性は、シナリオの最後の「不正の手法」のパートで操作される。古典的手法条件では、不正が架空の取引先を使った架空売上の計上という古典的手法でなされた旨の説明がなされる。他方、見積り・判断条件では、不正が会計上の見積り・判断に関してなされた旨（具体的には減損損失の未計上（将来の見積りのうち、非現実的かつ極めて楽観的な売上高やコスト削減の見積りをわざと採用））の説明がなされる。

被験者計画：今回は⁶、日本公認会計士協会近畿会 IT 部会と協力して、日本公認会計士協会の関西 3 支部（近畿会・京滋会・兵庫会）の会員・準会員 5,838 名（近畿会 4,289 名、京滋会 784 名、兵庫会 765 名）全員にメールにて web 実験への協力を要請した。実施期間は、2018 年 12 月から 2019 年 1 月の間である。各メールは 4 パターン（Web 実験のリンク URL が 4 パターン）あり、会員番号でランダムに 4 条件のうちどれか 1 つの条件を割当てることとし、被験者は、当該 UPL を辿り、どれか 1 つの条件だけに web から回答した。最終的なサンプル

⁶ 通常の陪審員実験では、被験者は一般人もしくは学生とされることが多いが（e.g. Grenier, Reffett, Simon, and Warne 2018）、今回は、以下のような 2 つの方針から、会計専門家である公認会計士および公認会計士試験合格者を実験の対象とした。①シナリオ自体が、会計の専門的な内容についてのものであること、②公認会計士の陪審員役としての意思決定と学生の陪審員役としての意思決定とを比較する Reffett, Brewster, and Ballou (2012)を参考に、今後、本研究を拡張する予定があること（まずファーストステップとして、公認会計士の意思決定データを採取し、その後、学生ないし一般人を被験者として同様のデータを採取する予定であること）。

ルサイズは、188 となった（平均年齢=48.78 歳、CPA =169 (準会員=19)、Big_N=128 (その他 60)、平均経験年数=17.35 年、監査法人の Partner 経験者= 71 (その他 117)）。

IV RESULTS

Summary statistics

前節で述べたように、各実験条件において、被験者は監査人の責任の度合いを「0: 全く責任がない」から「100: 全面的に責任がある」という 101 ポイントのリッカート・スケールで回答している。図表 3 は、各実験条件における当該尺度の平均値、標準偏差およびサンプルサイズを示している。

図表 3 記述統計

		AI 利用		
		なし (人のみ)	あり (AI 監査)	
見積り		80.69	81.97	81.29
		(17.37)	(25.18)	(19.16)
不正手法		[43]	[38]	[81]
		71.41	77.59	74.53
古典		(26.59)	(23.26)	(25.04)
		[53]	[54]	[107]
		75.57	79.40	
		(23.28)	(22.43)	
		[96]	[92]	

Note: 数値は、監査人の責任の度合い (0-100) の条件ごとの平均値。丸カッコ()内の数値は標準偏差、角括弧[]内の数値はサンプルサイズをそれぞれ示している。

図表 3 から、全体的傾向として、①AI 利用について、AI 利用がない場合よりも、AI 利用がある場合 (AI 監査) のほうが、監査人の責任はより重く判定されていること、②不正手法として、見積りの場合のほうが、古典的手法の場合よりも、監査人の責任はより重く判定されていることがわかる。

The result of ANOVA (hypothesis 1 and 2)

図表 4 は、AI 利用要因（利用あり vs. なし）と不正原因要因（古典的手法 vs. 見積り）を独立変数とする二要因の分散分析の結果を示している。分散分析の結果、不正原因の主効果は統計的に有意であった（ $F(1, 184) = 4.264, p < 0.05$ ）が、AI 利用の主効果（ $F(1, 184) = 1.337, p = 0.249$ ）、および、両要因の交互作用（ $F(1, 184) = 0.536, p = 0.465$ ）は統計的に有意ではなかった。

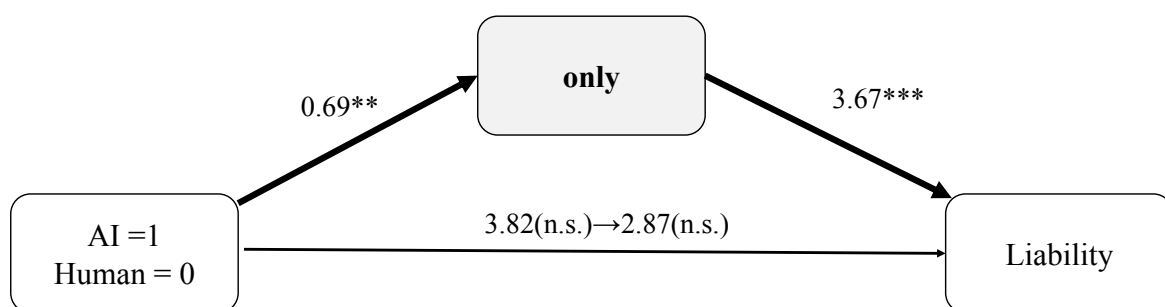
図表 4 2 要因分散分析

Source of Variation	SS	df	MS	F-statistic	p-value
AI 利用	689.00	1	688.86	1.337	0.249
不正原因	2,197.00	1	2197.38	4.264	0.040
AI 利用 × 不正原因	276.00	1	276.28	0.536	0.465
Residuals	94,812.00	184	515.28		

Mediation analysis for testing Hypothesis 1

仮説 1（AI 監査 vs 人のみ監査）について、被験者の心理プロセスをより深く理解するために、「AI の利用の有無」と「監査人の責任」との間の因果関係を、「意思決定の理由」変数を媒介とすることで説明できるか、媒介分析（Baron and Kenny 1986）を用いて検証する。特に、理由「監査の失敗の責任は、どんなときでも全面的に監査人にある」（「Only」変数と略す）に着目すると、図表 5 のようになる。

図表 5 H1 の媒介分析



Note: 図表は、H1 の媒介分析の結果を示している（ $N=188$ ）。左の「AI = 1, Human = 0」は、実験条件における「AI 利用の有無」を AI ダミー変数（利用あり=1, 利用なし = 0）で示している。右の「Liability」は、被験者が判断する監査人の責任の度合い（0-100）を示している。網掛けの「Only」は、監査人の責任の度合いと共に理由として被験者が「監査の失敗の責任

は、どんなときでも全面的に監査人にある」という問いに1-7の7段階リカートスケール（1: 全く賛成できない、7: 全く賛成できる）で解答した結果を表す。矢印の数値は係数、**は5%有意、***は1%有意、n.s.は有意差なしを示している。

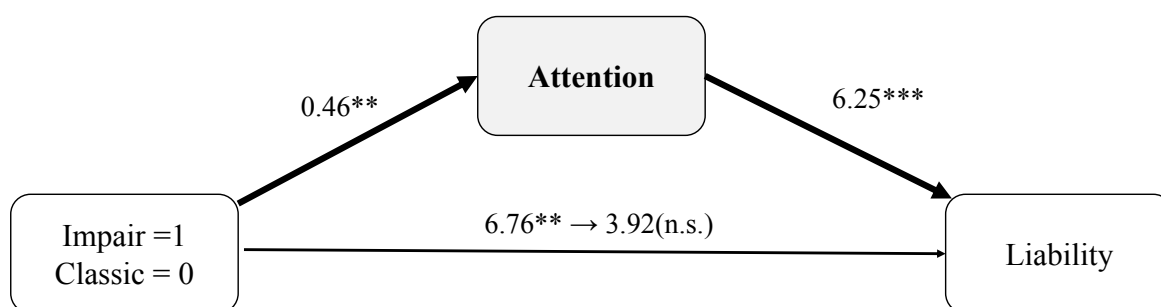
図表5に示されるとおり、「AIの利用の有無」と「監査人の責任」との間には、直接的に有意な効果は見られない。しかし、「Only」変数を媒介とすることで間接的に有意な効果が見られる。特に「AI」から「Only」へのパスがプラスに5%有意、さらに「Only」から「Liability」へのパスがプラスに1%有意であることは、次のように解することができる。すなわち、AI監査に対して（「AI=1」）、被験者は、たとえ監査人がAIを用いたとしても（責任はシェアされず）「監査人だけが責任を負わなければならない」と強く考え（「AI→only」プラスの効果）、そのことにより、「監査人の責任」をより大きく評価する（「only→Liability」プラスの効果）。このことから、媒介変数を用いると、仮説1は統計的に有意に支持される⁷。

Mediation analysis for preliminary testing Hypothesis 2

仮説2（古典的手法の不正 vs. 見積り不正）の予備的分析として、全条件における被験者の心理プロセスをより深く理解するために、「不正の原因」と「監査人の責任」との間の因果関係を、「意思決定の理由」変数を媒介とすることで説明できるか、媒介分析を用いて検証する。特に、理由「上記事例の会計不正は、正当な注意を払えば見抜けるレベルにある」（「Attention」変数と略す）に着目すると、図表6のようになる。

⁷ この媒介分析の結果は、AI (Decision aid)があることが、逆に人間に対する責任への注意を強く促し、人間の責任をより強く被験者に感じさせる(not share)ということを示している。このような人間心理は、行動経済学的にはレファレンス・ポイント (reference point) の議論などを用いて説明できるかもしれないが、この点の精緻な議論は別稿を期する予定である。

図表 6 H2 の媒介分析



Note: 図表は、H2 の媒介分析の結果を示している (N=188)。左の「Impair = 1, Classic = 0」は、実験条件における「不正の原因の違い」を Impair ダミー変数 (見積り不正 = 1, 古典的不正 = 0) で示している。右の「Liability」は、被験者が判断する監査人の責任の度合い (0-100) を示している。網掛けの「Attention」は、監査人の責任の度合いと共に理由として被験者が「上記事例の会計不正は、正当な注意を払えば見抜けるレベルにある」という問いに 1-7 の 7 段階リカートスケール (1: 全く賛成できない、7: 全く賛成できる) で解答した結果を表す。矢印の数値は係数、**は 5% 有意、***は 1% 有意、n.s. は有意差なしを示している。

図表 6 に示されるとおり、「不正の原因」と「監査人の責任」との間には、直接的に 5% 有意な効果がみられる。しかし、「Attention」変数を媒介とすると、その直接効果は消え、代わりに間接的に有意な効果が出現する。特に「Impair」から「Attention」へのパスがプラスに 5% 有意、さらに「Attention」から「Liability」へのパスがプラスに 1% 有意であることは、次のように解することができる。すなわち、会計上の見積りに係る不正 (「impair = 1」) に対して、被験者は、「監査人は正当な注意を払って当然見つけるべきであった」と強く考え (「Impair → Attention」プラスの効果)、そのことにより、「監査人の責任」をより大きく評価する (「Attention → Liability」プラスの効果)。このような間接的な効果の現れは、ある意味で、会計上の見積り不正に対する人間の役割に対する期待の強さを物語っている⁸。

ただし、上記の分析はフルサンプルによる (全ての条件に対する) 分析であり、他方、仮説 2 は AI 監査に限定したものであるため、次節では、AI 監査のサブサンプルに着目して仮説 2 を検証する。

⁸ ただし、このような結果は、被験者が、「正当な注意」という概念を知っている会計専門家 (公認会計士) であることに強く起因するのかもしれない。この点は、今後、(先に述べたような) 実験被験者の拡張を図る際に重要なポイントとなる。

Regression Analysis for testing hypothesis 2

上記の議論（古典的手法の不正 vs. 見積り不正）について、AI 監査のサブサンプルに限定して、「不正の原因」と「監査人の責任」との関係性を、その他の「意思決定の理由」変数を統制することで説明できるか、regression を用いて検証する。図表 7 は、被説明変数を「監査人の責任」とする重回帰分析の結果である⁹。

図表 7 H2 の Regression analysis

	Model 1	Model 2	Model 3	Model 4
	Estimate	Estimate	Estimate	Estimate
(Intercept)	29.09 ***	10.89	50.41 ***	25.47
impair_dummy	6.24	8.25 **	5.06	7.40 *
only	4.05 ***	2.32 *	4.23 ***	2.82 **
attention	5.30 ***	2.57	5.79 ***	3.16 *
maker		2.97		2.66
down		-1.26		-0.76
select		-0.07		-0.20
only2		5.47 ***		4.71 ***
maker2		0.85		0.95
AI_weak		1.60		1.98
AI_specialty		-3.15		-2.70
CPA			1.07	-0.57
Age			-0.57 **	-0.40
Big N			1.66	2.17
Experience			-0.02	-0.05
Partner			6.87	4.83
Obs.	92	92	92	92
adjusted R ²	0.321	0.396	0.359	0.408

Note: 図表は、AI 監査サンプルに限定した H2 の重回帰分析の結果を示している (N=92)。被説明変数は「Liability」（被験者が判断する監査人の責任の度合い (0-100)）である。「impair_dummy」は、不正原因が会計上の見積りであれば 1、古典

⁹ なお、被説明変数の「監査人の責任の度合い」の（数値に上限下限があるという）特性を鑑み Tobit model を用いた検証も行っているが、ほぼ同じ結果となることから、ここでは紙面の都合からその結果は割愛する。

的手法であれば 0 となる impair ダミー変数である。「Only」と「Attention」は、全ての条件において問うている「意思決定の理由」変数であり（Only：「監査の失敗の責任は、どんなときでも全面的に監査人にある」、Attention：「上記事例の会計不正は、正当な注意を払えば見抜けるレベルにある」）、「maker」から「AI_specialty」は、AI 監査条件にのみ問うている「意思決定の理由」変数であり、全て、被験者が 1-7 の 7 段階リカートスケール（1: 全く賛成できない、7: 全く賛成できる）で回答した変数である（具体的な設問は Appendix 1 参照）。「CPA」から「Partner」は、被験者の個人特性を示すデモグラフィックデータである（具体的内容は Appendix 2 参照）。各数値は係数、*は 10%有意、**は 5%有意、***は 1%有意を示している。

図表 7 に示されるとおり、AI 監査に限定したサブサンプルにおいて、意思決定理由や個人特性を制御すると、「不正の原因」を示すダミー変数 (impair dummy) は 5%水準(Model 2) もしくは 10%水準(Model 4) でプラスに有意となることから、会計上の見積りを原因とする不正に対して、被験者は監査人の責任をより高く判定することが示唆され、仮説 2 は統計的に有意に支持される。

V DISCUSSION AND CONCLUSION

本研究は、監査人が監査業務に AI を利用することが、監査人の不正を見逃した場合の法的責任にどのように影響を与えるか、また不正の多様性（不正の原因の違い）がどのような影響を与えるかについて、日本の公認会計士 188 名を被験者に、「AI 利用の有無」（利用あり vs. 利用なし）と「不正の原因」（古典的不正 vs. 会計上の見積り判断による不正）とを考慮した 2×2 between-participant design の場面想定法を用いた陪審員の意思決定実験により検証した。本研究の結果をまとめると、図表 8 になる。

図表 8 本研究の結果：まとめ

<p>論点 1・AI 監査 vs 従来の監査（監査の失敗）→AI を利用することで、人の責任がより強調されてしまう 【意図せざる帰結】</p>	<p>AI 監査…仮に監査の質向上&「監査の失敗」が減る可能性があるとしても、</p>
<p>論点 2・AI 監査：不正原因の多様性（古典的不正 vs 見積り判断不正） → 見積り判断不正を見逃すことで人の責任はより重くなる 【見積り判断への対処=人がやるべき業務 (not AI)】</p>	<p>不正を見逃した際の監査人の責任は、これまで以上に大きくなるおそれ</p>

図表 8 に示されるとおり、①AI の利用は、監査人の責任をより重くする恐れがあること、また、②特に AI を利用する場合、会計上の見積り判断による不正を見逃した際に、監査人の責任はより重くなる可能性があることが、本研究から示唆される。①は、AI を利用した

監査の逆効果といえる。また②は、AI 監査において、監査人は今まで以上に見積り判断など人間が得意とする（AI が不得意とする）項目に注力した監査を進める必要があることを示唆する。

本研究は、「会計不正の多様性」と AI 監査の責任問題に接近した数少ない実験研究であるが、ここでの仮説検証から得られるインプリケーションは次の2つである。まず第1に、AI 時代の監査のあり方に係る実務的な示唆である。本研究の結果から、AI 監査における監査人の責任はより重くなってしまふことが理解できる。このことは逆に、AI 監査において、監査法人が社会に対して積極的に情報発信することの重要性を示唆している。たとえば、現状では、AI 監査によって、監査の質が向上することや、不正の量が減少することは、あまり社会に向けてアピールされていない。これらを社会に発信するだけでなく、制度的にも監査における AI 利用のガイドラインを整備・明確化し、監査において、どのような点で AI が利用され、またどのような点で人間の監査人が判断をおこなうのかを、社会に伝える制度的努力が求められる。

また、第2は、AI の社会的受容（social acceptance）の理論に対する示唆である。AI の社会的受容については、実験社会科学や科学技術社会論などで様々な議論がなされている。たとえば、自動運転の倫理ジレンマ研究の成果（Awad et al. 2018; Bonnefon et al. 2016）によれば、人間は、AI の意思決定が（人間の）既存の社会ルールに沿ったかたちでなされるべきであるという無意識のバイアスを有していることが明らかにされている。これに対して、本研究は、AI を人間が利用する場合の人間のあり方に焦点をおく点で、これまでの先行研究とは違う視点から社会的受容の問題を捉えている。特に本研究は、AI を利用することで、人間のなすべき役割の重要性が減じるのではなく、むしろ高まるということを示しており、この点は、AI と人間とが共存する社会を構築するうえでの重要な示唆といえる。特に現在は、COVID-19 問題を背景として企業の DX(Digital Transformation)化が進んでいるが、本稿はテクノロジーと経済社会の共存のあり方の議論に対しても一定の示唆を与える。

最後に、本研究の限界と今後の展望として、被験者計画について述べる。通常の陪審員実験では、被験者は一般人もしくは学生とされることが多い（e.g. Grenier, Reffett, Simon, and Warne 2018）。これに対して今回は、脚注6で述べたとおり、以下のような2つの方針から、会計専門家である公認会計士および公認会計士試験合格者を実験の対象とした。すなわち、①シナリオ自体が、企業会計の専門的な内容についてのものであること、②公認会計士の陪審員役としての意思決定と学生の陪審員役としての意思決定との違いを比較する Reffett, Brewster, and Ballou (2012)を参考に、今後、本研究を拡張する予定があることである。よって、

本研究をベースにして、学生ないし一般人を被験者とした追加の陪審員実験をおこなうことで、AI 監査のあり方をより深く検証していくこととしたい。これは今後の課題である¹⁰。

付記：本稿は、JSPS 科研費 JP 19K21710, 19K21711、および「同志社大学新型コロナウイルス感染症に関する緊急研究課題プロジェクト助成金」の助成を受けた研究の一部である。

¹⁰ AI 時代の会計・監査の諸課題については、Gonzalez and Hoffman (2018)のほか、田口 (2018b; 2019)なども併せて参照。

Appendix 1 「意思決定の理由」変数

実験条件	No	変数名	被験者への質問項目（被験者は各項目への同意度を 1-7 の 7 段階で回答： 1:全く同意できない、7:全く賛成できる）
全 4 条件 共通	1	only	監査の失敗の責任は、どんなときでも全面的に監査人にある
	2	attention	上記事例の会計不正は、正当な注意を払えば見抜けるレベルにある
AI 利用 あり条件 のみ	3	maker	上記事例の会計不正は、AI が見抜くべきものであるから、AI 作成者側の責任である
	4	down	上記事例の監査の失敗は AI がおこなった結果なので、監査人の責任は軽減されるべきである
	5	select	監査人は AI を利用することを選んだ責任を負うため、監査人の責任はより重くなる
	6	only2	AI の利用にかかわらず、監査人は「監査の失敗」の責任を負うべきである
	7	maker2	AI を提供した企業は製造物責任を負うべきである
	8	AI_weak	AI には得意な業務と不得意な業務があり、今回の事例では監査人の責任は重い
	9	AI_specialty	AI には得意な業務と不得意な業務があり、今回の事例では監査人の責任は軽い

Appendix 2. 被験者の個人特性を示すデモグラフィックデータ

変数名	内容
CPA	日本公認会計士協会の会員(CPA) = 1, 準会員(資格合格者) = 0 とするダミー変数
Age	被験者の年齢
Big N	大手監査法人所属 =1, その他所属 =0 とするダミー変数
Experience	被験者の実務経験年数
Partner	監査法人の社員経験あり = 1, なし = 0 とするダミー変数

References

- Autor, D. H. and Dorn, D. 2013. The growth of low skill service jobs and the polarization of the U.S. labor market. *American Economic Review* 103(5):1553-1597.
- Awad, E., S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J. F. Bonnefon, and I. Rahwan. 2018. The Moral Machine experiment. *Nature* 563:59-64.
- Baron, R. M. and D. A. Kenny. 1986. The moderator? mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51: 1173-1182.
- Bonnefon, J. F., A. Shariff, and I. Rahwan. 2015. The social dilemma of autonomous vehicles. *Science* 352(6293):1573-1576.
- Brasel, K., M. M. Doxey, J. H. Grenier, and A. Reffett. 2016. Risk Disclosure Preceding Negative Outcomes: The Effects of Reporting Critical Audit Matters on Judgments of Auditor Liability. *The Accounting Review* 91(5): 1345-1362.
- Frey, C. B. and M. A. Osborne. 2017. The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change* 114: 254–280.
- Gimbar, C., B. Hansen, and M. E. Ozlanski. 2016. The Effects of Critical Audit Matter Paragraphs and Accounting Standard Precision on Auditor Liability. *The Accounting Review* 91(6): 1629-1646.
- Grenier, J. H., D. J. Lowe, A. Reffett and R. C. Warne. 2015. The Effects of Independent Expert Recommendations on Juror Judgments of Auditor Negligence. *AUDITING: A Journal of Practice & Theory* 34(4); 157-170.
- Grenier, J. H., B. Pomeroy., and M. T. Stern. 2015. The Effects of Accounting Standard Precision, Auditor Task Expertise, and Judgment Frameworks on Audit Firm Litigation Exposure. *Contemporary Accounting Research* 32(1): 336-357.
- Grenier, J.H., A. Reffett, C. A. Simon, and R. C. Warne. 2018. Researching Juror Judgment and Decision Making in Cases of Alleged Auditor Negligence: A Toolkit for New Scholars. *Behavioral Research in Accounting* 30(1): 99-110.
- Gonzalez, G. C., and V. B. Hoffman 2018. Continuous Auditing's Effectiveness as a Fraud Deterrent. *AUDITING: A Journal of Practice & Theory* 37(2): 225-247.
- Jackson, S. 2008. The effect of firms' depreciation method choice on managers' capital investment decisions. *The Accounting Review* 83 (2): 351–376.
- Jackson, S., T. Rodgers, and B. Tuttle. 2010. The effect of depreciation method choice on asset selling prices. *Accounting, Organizations and Society* 35 (8): 757–774.

- Kadous, K. and M. Mercer. 2012. Can Reporting Norms Create a Safe Harbor? Jury Verdicts against Auditors under Precise and Imprecise Accounting Standards. *The Accounting Review* 87(2): 565-587.
- Kadous, K. and M. Mercer. 2016. Are Juries More Likely to Second-Guess Auditors under Imprecise Accounting Standards?. *AUDITING: A Journal of Practice & Theory* 35(1): 101-117.
- Lowe, D. J., P. M. J. Reckers, and S. M. Whitecotton. 2002. The Effects of Decision - Aid Use and Reliability on Jurors' Evaluations of Auditor Liability. *The Accounting Review* 77(1):185-202.
- McAfee, A., and E. Brynjolfsson. 2018. *Machine, Platform, Crowd: Harnessing Our Digital Future*. WW Norton & Co Inc (マカフィー・ブリニョルフソン(村井章子訳)2018. 『プラットフォームの経済学』日経 BP 社)
- Murdoch, T. B. and A. S. Detsky 2013. The Inevitable Application of Big Data to Health Care. *Journal of the American Medical Association* 309(13): 1351-1352.
- Rissland, E. L., K. D. Ashley, and R. P. Loui 2003. AI and Law: A fruitful synergy. *Artificial Intelligence*. 150:1-15.
- Reffett, A., B. E. Brewster, and B. Ballou. 2012. Comparing auditor versus nonauditor assessments of auditor liability: An experimental investigation of experts' versus lay evaluators' judgments. *AUDITING: A Journal of Practice & Theory* 31 (3): 125-148.
- Seybert, N. 2010. R&D capitalization and overinvestment in continuing projects. *The Accounting Review* 85 (2): 671-693.
- 田口聡志 2015. 『実験制度会計論 -未来の会計をデザインする--』中央経済社
- 田口聡志 2018a. 「人間と AI とが共存する未来社会のデザイン：実験社会科学、トランス・サイエンス、フューチャー・デザインの融合へ向けて」 『同志社商学』 69(6)、177-202.
- 田口聡志 2018b. 「AI時代の監査報酬を考える -A preliminary report-」 日本監査研究学会課題別研究部会編 『テクノロジーの進化と監査(平成 30 年度中間報告)』 第 12 章、120-145.
- Taguchi, S. 2018. An experimental study on the Social Dilemma on legal responsibility when accounting professionals are replaced with Artificial Intelligence. *mimeo*.
- 田口聡志 2019. 「AI 時代の会計の質の変容と『フューチャー・ハザード』」 『企業会計』 71(1) : 89-96.