

# 会計監査×AI・機械学習・NLPの研究動向：10年間の学術的發展（2015-2025）

会計監査分野におけるAI・機械学習・自然言語処理の研究は、2015年頃の従来型機械学習から、2025年現在の大規模言語モデル（LLM）活用へと劇的な進化を遂げた。**従来の統計的手法を用いた不正検出精度60-70%台から、深層学習により90%超の精度が達成可能**となり、監査実務への実装が急速に進んでいる。特に、テキストマイニングの分野では、Loughran-McDonald辞書による単純な感情分析から、FinBERTやGPTベースのモデルへと発展し、財務諸表の言語的特徴から不正リスクを高精度で予測できるようになった。本レビューでは、査読付き学術論文を中心に、この10年間の研究動向を網羅的に整理する。

## 第1期：基盤構築と機械学習の本格導入（2015-2017年）

### 不正検出研究の黎明期

2015年頃の会計監査AI研究は、従来の統計的手法に機械学習を組み合わせるアプローチが主流であった。この時期の最も影響力のある研究は、\*\*Perols（2011年、Auditing: A Journal of Practice & Theory）\*\*による不正検出モデルの比較研究である。同研究は、ロジスティック回帰、SVM、ニューラルネットワーク、バギング、C4.5、スタッキングを比較し、**監査人の交代、裁量的発生高、Big4監査人、売掛金比率**などが不正予測の重要な変数であることを特定した。被引用数500件以上を記録し、この分野の基盤的論文となった。

\*\*Hajek & Henriques（2017年、Knowledge-Based Systems）\*\*は、財務比率とテキスト特徴量を組み合わせた不正検出モデルを提案し、**機械学習が人間の監査人が見逃すパターンを捉えられる**ことを実証した。決定木、SVM、ランダムフォレスト、ナイーブベイズを比較し、収益性指標とテキストセンチメントが主要予測因子であることを明らかにした。

\*\*Perols, Bowen, Zimmermann & Samba（2017年、The Accounting Review）\*\*は、トップ会計学ジャーナルに掲載された重要論文であり、データ分析の前処理手法（特徴選択、インスタンス選択、アンサンブル手法）を導入し、**不正予測精度を約10%向上**させた。

### 監査自動化の理論的枠組み

監査プロセスの自動化に関する研究は、\*\*Issa, Sun & Vasarhelyi（2016年、Journal of Emerging Technologies in Accounting）\*\*によるAI監査の研究アジェンダ提案から本格化した。監査業務が労働集約的であり、意思決定構造が多様であることから、部分的自動化に適していると主張し、深層学習、機械学習、エキスパートシステムの統合可能性を示した。

\*\*Kokina & Davenport（2017年、JETA）\*\*は、Big4会計事務所によるAI導入状況を調査し、認知技術がルールベースタスクの自動化を可能にすることを報告した。\*\*Appelbaum, Kogan & Vasarhelyi（2017年、Auditing: A Journal of Practice & Theory）\*\*は、ビッグデータ時代における監査エンゲージメントの変革を論じ、従来のサンプリングがビッグデータ環境では意味を失いつつあると指摘した。

### NLP・テキストマイニングの辞書ベースアプローチ

この時期のNLP研究は、**辞書ベースの感情分析**が主流であった。\*\*Loughran & McDonald（2016年、Journal of Accounting Research）\*\*は、財務テキスト分析のベストプラクティスを確立したサーベイ論文であり、一般的な辞

書が財務文脈では不適切であること（Harvard IV辞書は楽観語の83%、悲観語の70%を誤分類）を指摘し、ドメイン特化型アプローチの重要性を強調した。

**\*\*Dyer, Lang & Stice-Lawrence（2017年、Journal of Accounting and Economics）\*\*は、LDA（Latent Dirichlet Allocation）トピックモデリングを用いて10-Kファイリングの進化を分析した。1996年から2013年の10-Kを150トピックに分類し、開示の長さ、定型文、冗長性が増加する一方、具体性と可読性が低下していることを発見した。**

---

## 第2期：深層学習の台頭とRPA統合（2018-2020年）

### 機械学習による不正検出の飛躍

2018年から2020年は、**アンサンブル学習と深層学習が不正検出精度を劇的に向上**させた転換期である。

**\*\*Bao, Ke, Li, Yu & Zhang（2020年、Journal of Accounting Research）\*\*は、この分野の画期的論文であり、\*\*財務比率ではなく生の会計数値を用いたアンサンブル学習（Random Forest, XGBoost）\*\*が、従来のDechow et al.ロジスティック回帰モデルやCecchini et al.SVMモデルを大幅に上回ることを実証した。28の生会計変数と14の財務比率を使用し、新しい評価指標NDCG@kを導入した。データとコードを公開し、再現可能な研究の模範となった。**

**\*\*Craja, Kim & Lessmann（2020年、Decision Support Systems）\*\*は、\*\*Hierarchical Attention Network（HAN）\*\*を用いて財務比率とMD&Aテキストを統合した不正検出モデルを提案した。文書構造を捉え、単語・文レベルで不正の「レッドフラッグ」を識別できることを示した深層学習応用の先駆的研究である。**

**\*\*Dong, Liao & Zhang（2018年、Journal of Management Information Systems）\*\*は、金融ソーシャルメディアデータを活用した不正検出を提案し、財務諸表以外の非構造化データの価値を実証した。**

### RPA（ロボティック・プロセス・オートメーション）の監査適用

**\*\*Moffitt, Rozario & Vasarhelyi（2018年、JETA）\*\*は、RPAが監査モデルを根本的に変革する可能性を論じ、ルールベースで反復的な手作業タスクの自動化により、監査人がより高次の思考に集中できるようになると主張した。**

**\*\*Huang & Vasarhelyi（2019年、International Journal of Accounting Information Systems）\*\*は、4段階のRPAフレームワークを提案し、確認プロセスの自動化パイロットプロジェクトを成功させた。確認手続きの大幅な時間短縮と精度向上を実証した。**

### 継続監査システムの発展

**\*\*Dai & Vasarhelyi（2020年、JETA）\*\*は、\*\*Continuous Audit Intelligence as a Service（CAIaaS）\*\*アーキテクチャを提案し、経験の浅い監査人でも高度な技術を活用できるクラウドベースのインテリジェントアプリ推奨システムを構想した。**

### NLPのWord Embedding時代への移行

**\*\*Yang, Uy & Huang（2020年、arXiv）\*\*は、FinBERTの初期バージョンを発表した。4.9億トークン（10-K/10-Q、アナリストレポート、決算説明会）で事前学習されたBERTモデルは、汎用BERTを大幅に上回る財務感情分類精度を達成した。**

**\*\*Campbell, Lee, Lu & Steele（2020年、Contemporary Accounting Research）\*\*は、決算説明会と10-Kファイリングの開示トーンの時間変動を分析し、投資家が経営者のコミュニケーションパターンの変化から学習すること**

を発見した。

## LSTM・RNNの金融時系列への応用

**\*\*Fischer & Krauss (2018年、European Journal of Operational Research) \*\***は、**\*\*LSTM (Long Short-Term Memory) \*\***の大規模金融市場予測への初適用を報告し、高ボラティリティ、平均以下のモメンタム、極端な方向性変動が予測力の源泉であることを特定した。

**\*\*Chi & Chu (2021年、Sustainability) \*\***は、LSTMとRNNを台湾企業の不正検出に適用し、**LSTMがRNNを全指標で上回る**ことを実証。14の財務変数と4のコーポレートガバナンス変数を使用した。

---

## 第3期：Transformer・LLMと説明可能AI（2021-2025年）

### FinBERTと金融NLPの革新

**\*\*Huang, Wang & Yang (2023年、Contemporary Accounting Research) \*\***は、FinBERTの包括的評価を発表し、**感情分類精度88.2%**（L-M辞書62.1%、従来ML71-74%に対して）を達成した。金融固有語彙（一般テキストでは稀だが金融テキストで頻出する単語）がFinBERTの優位性を駆動し、他手法が中立と分類する曖昧な文の感情識別に優れることを示した。

**\*\*Bochkay, Chychyla, Leone & Ferreira (2023年、Contemporary Accounting Research) \*\***は、会計分野のテキスト分析手法を体系化し、単純変換、テキスト比較、従来ML、深層学習の4カテゴリに分類した枠組みを提案。深層学習とTransformerモデルの活用拡大を提唱した。

**\*\*Ergun (2025年、Intelligent Systems in Accounting, Finance and Management) \*\***は、**GPTベースのLLMによる財務諸表不正検出**を報告し、テキストパターン分析と財務比率の統合により高精度予測を実現した。

### 説明可能AI（XAI）の監査への導入

**\*\*Zhang, Cho & Vasarhelyi (2022年、International Journal of Accounting Information Systems) \*\***は、**監査におけるXAIの基盤的論文**であり、LIME（Local Interpretable Model-agnostic Explanations）とSHAP（Shapley Additive exPlanations）を重要な修正再表示リスク予測に適用した。PCAOB監査証拠基準（AS 1105）および文書化基準（AS 1215）との関連を明示し、規制コンプライアンスにおける解釈可能性の必要性を論じた。

**\*\*Zhong et al. (2024年、Current Issues in Auditing) \*\***は、不正検出における透明性確保のためのXAI統合を実証し、モデル精度と透明性のトレードオフ、新興AI規制への対応を議論した。

### 継続企業の前提予測における深層学習

アジア太平洋地域、特に**台湾**からの研究が継続企業予測分野をリードしている。

**\*\*Chi & Shen (2022年、Sustainability) \*\***は、2段階ハイブリッドアプローチ（CART/CHAIDによる変数選択 → XGB, ANN, SVM, C5.0による分類）を提案し、**CHAID-C5.0で95.65%の精度**を達成した。536社の台湾上場企業（2000-2019年）を分析し、負債比率、ROE、ROA、EPSが主要予測因子であることを特定した。

**\*\*Jan (2021年、Information) \*\***は、**CART-RNNハイブリッドモデル**で95.28%の精度を達成し、RNNのGoing Concern予測への初適用を報告した。**\*\*Chi & Chu (2021年、Sustainability) \*\***は、**LSTMで96.15%の精度**を達成し、深層学習の統計的手法に対する優位性を実証した。

### プロセスマイニングと内部統制評価

**\*\*Duan, Vasarhelyi & Codesso (2025年、Journal of Information Systems) \*\*は、プロセスマイニングと機械学習を統合した内部統制評価モデル**を提案した。4コンポーネント（プロセスマイニング分析、ルールベース変異分析、統制リスク評価、異常検出）からなるモデルは、統制の弱点と欠落を特定し、高リスク領域への調査を誘導する。

**\*\*Nonnenmacher & Gomez (2021年、IJAIS) \*\*は、財務諸表監査へのプロセスマイニング埋め込みの実現可能性**を実証し、手動監査手続きを自動化されたプロセス分析で置き換えることで、監査結論の信頼性が向上することを示した。

## ブロックチェーンと監査の統合

**\*\*Cao et al. (2024年、Management Science) \*\*は、許可型ブロックチェーンと安全なマルチパーティ計算による財務報告・監査の革新を論じ、監査事務所のコスト70%削減**の可能性を示した。プライバシーと検証の課題に対処しつつ、自動化された継続的監査を実現するフレームワークを提案した。

## AIの倫理・ガバナンス

**\*\*Laine & Mäntymäki (2024年、Information & Management) \*\*は、倫理ベースのAI監査に関する体系的文献レビュー**を実施し、公平性、透明性、非有害性、責任、プライバシー、信頼、善行、自律性の8つの倫理原則を統合した。

**\*\*Mökander et al. (2023年、AI and Ethics) \*\*は、LLM監査の3層フレームワーク**を提案した：

- 第1層：ガバナンス監査（組織手続き、説明責任）
- 第2層：モデル監査（訓練後、デプロイ前の能力評価）
- 第3層：アプリケーション監査（下流利用における倫理的整合性）

---

## 日本・アジア太平洋地域からの研究貢献

### 日本発の不正検出研究

**\*\*Yazawa, Araragi, Itakura et al. (2024年、SSRN) \*\*は、\*\*NLPと加重ランダムフォレスト (Weighted Random Forest) \*\*を用いた日本上場企業の財務不正検出を報告した。2010-2019年のデータを分析し、**AUC 0.907**を達成。不正企業のMD&Aがネガティブなトーン、複雑な文体、比率ベース表現の少なさを示すことを発見した。**

日本の**\*\*産業技術総合研究所 (AIST) \*\*は、2024年に機械学習品質管理ガイドライン**を発表し、利用時品質、外部品質、内部品質の分類枠組みと品質管理手法を提示した。

### 台湾・中国からの継続企業予測研究

台湾は継続企業予測研究の中心地であり、前述のChi、Jan、Gooらの研究群が世界最高水準の予測精度を達成している。**Yeh, Chi & Lin (2014年、Information Sciences) は、ランダムフォレストとRough Set理論のハイブリッドで96.10%の精度**を報告した。

中国からは、**LightGBMとSHAP解釈可能性**を組み合わせた財務困難予測（2024年、Pacific-Basin Finance Journal）や、**XGBoostとNSGA-II最適化**による解釈可能な早期警戒フレームワーク（2025年、Economic Modelling）など、機械学習とXAIを統合した研究が増加している。

技術進化の全体像

時期	主要技術	主要応用領域	代表的精度
2015-2017	ロジスティック回帰、SVM、決定木、LDA	不正検出、テキスト感情分析	70-80%
2018-2020	アンサンブル学習（XGBoost, RF）、LSTM、Word2Vec	不正予測、継続企業予測、RPA	85-90%
2021-2023	BERT/FinBERT、深層ニューラルネット、SHAP/LIME	NLP不正検出、XAI、監査品質予測	88-96%
2024-2025	GPT/LLM、Graph Transformer、マルチモーダル分析	LLM監査、ブロックチェーン統合、倫理監査	90%超

主要学術ジャーナルと研究拠点

トップ会計学ジャーナル

- **The Accounting Review** (Perols et al. 2017)
- **Journal of Accounting Research** (Bao et al. 2020; Loughran & McDonald 2016)
- **Contemporary Accounting Research** (Huang et al. 2023 FinBERT)
- **Auditing: A Journal of Practice & Theory** (Appelbaum et al. 2017)
- **Review of Accounting Studies** (Bertomeu et al. 2021)

情報システム・意思決定支援

- **Journal of Management Information Systems** (Dong et al. 2018)
- **International Journal of Accounting Information Systems** (Zhang et al. 2022 XAI)
- **Decision Support Systems** (Craja et al. 2020)
- **Journal of Emerging Technologies in Accounting** (Issa et al. 2016; Moffitt et al. 2018)

AI・機械学習特化

- **Expert Systems with Applications**
- **Knowledge-Based Systems** (Hajek & Henriques 2017)
- **European Journal of Operational Research** (Fischer & Krauss 2018)

結論：研究動向の要点と今後の方向性

この10年間の研究は、会計監査におけるAI・機械学習・NLPの応用が**理論的探索から実用的実装へ**と移行したことを示している。

**技術的には**、辞書ベースのテキスト分析からTransformer/LLMへ、単一分類器からアンサンブル・深層学習へ、ブラックボックスモデルから説明可能AIへという3つの明確な進化軸が確認された。**精度面では**、Bao et al. (2020) が示したように、生の会計数値と高度な機械学習の組み合わせが従来の財務比率ベースモデルを大幅に上回り、Chi & Chu (2021) のLSTMは継続企業予測で96%超の精度を達成した。

**地域的には**、米国が不正検出とNLP研究をリードする一方、**台湾が継続企業予測で世界最高水準の研究**を産出し、中国からのXAI統合研究、日本からのNLP応用研究が増加している。

**今後の研究課題**として、国際的なモデル検証、リアルタイム予測システムの開発、LLM統合の深化、非財務変数（ESG、ガバナンス）の活用拡大、そして規制コンプライアンスに対応した説明可能性の強化が挙げられる。会計監査のAI化は、もはや「可能かどうか」ではなく「いかに実装するか」の段階に入っている。