

2026 年度夏期
グラデュエーションペーパー
予稿

題 目	
小売業における生成 AI 導入の財務効果と現場実態	
技術経営論文	ビジネス企画提案

学籍 番号	8824102	氏 名	大浦 直也
----------	---------	--------	-------

教 員	
主 査	井上 悟志 教授
審査 委員 担当	田村 浩道 教授

「小売業における生成 AI 導入の財務効果と現場実態」

目次

第1章 序論	7
1.1 研究背景	7
1.1.1 日本企業の IT 投資と生成 AI への期待	7
1.1.2 生成 AI 投資の「現実」	7
1.1.3 期待と現実のギャップ	8
1.2 研究目的	8
1.3 研究仮説	9
1.4 本研究の意義	9
1.5 論文構成	10
第2章 先行研究と理論的枠組み	11
2.1 AI・DX 投資と企業業績に関する先行研究	11
2.1.1 DX 投資の財務的効果	11
2.1.2 テキスト分析による企業開示研究	11
2.1.3 シグナリング理論と情報開示	11
2.1.4 投資コスト仮説	12
2.2 小売業における DX・生成 AI 活用の実態	12
2.2.1 小売業の DX 化の現状	12
2.2.2 生成 AI の小売業への適用可能性	12
2.3 小売業に関する定量研究の限界と本研究の位置づけ	12
2.4 理論的枠組みのまとめ	13
第3章 データと変数	14
3.1 データソース	14
3.1.1 EDINET（有価証券報告書テキスト）	14
3.1.2 J-Quants（財務データ・業種分類）	14
3.2 サンプルの概要	15
3.3 キーワード変数の定義	15
3.3.1 キーワードカテゴリと重み	15

3.3.2 スコア算出方法	16
3.3.3 キーワードスコアの記述統計	16
3.4 財務変数の定義	17
3.4.1 目的変数	17
3.4.2 コントロール変数	17
3.4.3 逆因果コントロール変数	17
3.5 分析設計の概要	18
第4章 分析手法	19
4.1 分析の全体設計	19
4.1.1 分析の基本構造：1期ラグ設計	19
4.1.2 研究仮説と分析手法の対応	19
4.1.3 分析の階層構造	20
4.2 分析手法の選択根拠	22
4.2.1 OLS 回帰の基本的性質	22
4.2.2 代替手法との比較による選択基準	23
4.2.3 1期ラグを採用した根拠	23
4.3 グループ比較分析	24
4.3.1 四分位分類によるグループ比較	24
4.3.2 統計的検定	24
4.4 ラグ OLS 回帰分析（メイン分析）	25
4.4.1 基本モデルの定式化	25
4.4.2 コントロール変数の選択根拠	25
4.4.3 多仕様変数探索（140仕様）	26
4.4.4 OLS 回帰の前提条件と本研究での対処	26
4.5 逆因果検証（前期業績コントロール）	27
4.5.1 逆因果の構造と問題	27
4.5.2 前期業績コントロールモデル	27
4.6 パネル固定効果モデル（ロバスト性検証）	28
4.6.1 固定効果モデルが解決する問題	28

4.6.2 分析対象の選択根拠	29
4.7 差分の差分法 (DiD)	29
4.7.1 分析の設計	29
4.7.2 DiD の限界と位置づけ	29
4.8 業種横断スキャン (探索的分析)	30
4.8.1 分析設計	30
4.8.2 多重比較への対処	30
4.9 分析手法の限界と補完戦略	31
第 5 章 分析結果	32
5.1 知見①: 全業種において dxScore は翌期 ROA と正に相関する	32
5.1.1 主張	32
5.1.2 ラグ OLS 回帰の結果	32
5.1.3 ロバスト性の評価	33
5.1.4 解釈上の注意点	34
5.2 知見②: 情報通信業において aiScore は翌期営業利益率・純利益率と負に相関する	34
5.2.1 主張	34
5.2.2 ラグ OLS 回帰の結果	34
5.2.3 ロバスト性の評価	35
5.2.4 解釈上の注意点	36
5.3 知見③: 小売業においてキーワードスコアと業績に有意な関係は存在しない	36
5.3.1 主張	36
5.3.2 分析結果	36
5.3.3 解釈	37
5.4 知見④: DX/AI 言及が翌期業績に正の効果をもたらす業種はガラス・土石製品と非鉄金属に限定される	38
5.4.1 主張	38
5.4.2 正の効果が有意な業種	38
5.4.3 負の効果が有意な業種	38
5.4.4 業種別効果の構造	39

5.5 知見⑤：全業種合計の「dxScore → ROA 正効果」は業種内傾向ではなく合成効果	39
5.5.1 主張	39
5.5.2 統計的根拠	39
5.5.3 解釈	40
5.6 分析結果のまとめ	40
第 6 章 考察	43
6.1 全業種における dxScore → ROA 正効果の解釈（知見①）	43
6.1.1 シグナリング仮説との整合性	43
6.1.2 シンプソンのパラドックス的構造（知見⑤）	43
6.2 情報通信業における投資コスト先行パターン（知見②）	43
6.2.1 AI 投資タイムラグ仮説	43
6.2.2 dxScore と aiScore の異なる性格	44
6.2.3 中長期的な回復可能性	44
6.3 小売業の特殊性とデータ上の限界（知見③）	44
6.3.1 小売業の AI 開示水準の構造的低さ	44
6.3.2 キーワード辞書の限界	45
6.4 素材製造業（グループ A）の解釈（知見④）	45
6.5 先行研究との対話	45
第 7 章 アンケート調査：小売業従事者の生成 AI 活用実態と効果認識	46
7.1 調査概要	46
7.1.1 調査の位置づけと目的	46
7.1.2 調査設計	46
7.1.3 設問構成	46
7.2 回答者の基本属性	47
7.2.1 業種別内訳	47
7.2.2 分析上の群分け	47
7.3 生成 AI 効果測定の実態（Q7）	48
7.4 生成 AI の各観点への寄与評価（Q8）	48
7.5 統計的検証	49

7.5.1 仮説設定	49
7.5.2 H5-1：効果測定実施群 vs 未測定群（Welch の t 検定）	49
7.5.3 H5-2：Q8 観点間の相関分析（Pearson 相関）	50
7.5.4 H5-3：評価理由の群間比較（ χ^2 検定）	50
7.5.5 H5-4～H5-6：本社部門 vs 店舗部門の比較	51
7.6 考察	52
7.6.1 効果測定の不在がもたらす AI 評価の希薄化	52
7.6.2 「AI 総合印象」として現れる評価構造	53
7.6.3 部門差の非対称性：個人軸のみ本社優位	53
7.6.4 仮説 H5 の検証結果	53
7.6.5 H5-1～H5-6 の検証結果まとめ	54
第 8 章 結論・限界・今後の課題	55
8.1 研究のまとめ	55
8.1.1 研究仮説の検証結果	55
8.1.2 仮説の検証結果	56
8.2 学術的貢献	57
8.3 実務的含意	57
8.4 本研究の限界	58
8.4.1 観察期間の短さ	58
8.4.2 genAiScore の非有意	59
8.4.3 キーワード辞書の限界	59
8.4.4 因果推論の限界	59
8.4.5 アンケート調査の限界	59
8.5 今後の研究課題	59
8.5.1 長期追跡研究	59
8.5.2 業種別専用辞書の開発	59
8.5.3 グループ A（素材製造業）の深掘り	60
8.5.4 プロセス産業ダミーの設計	60
8.5.5 現場と経営開示の連動分析	60

8.6 海外の動向との対比：NRF Big Show 2025・2026 の知見	60
8.6.1 NRF Big Show の概要	60
8.6.2 NRF 2025 の主要報告	60
8.6.3 NRF 2026 の知見：「実験から実装へ」の転換	61
8.6.4 報告の性格と留意点	63
8.6.5 日本の小売業との対比	63
参考文献	65
付録 66	
Appendix A：スクリーニングアンケート設問一覧	66
概要 66	
設問一覧と回答結果	66
Appendix B：本調査アンケート設問一覧	68
概要 68	
設問一覧と回答結果	68

研究背景と問題意識

なぜ今、この研究が必要か

2022 年末の ChatGPT 公開以降、生成 AI は DX 推進の有力な施策として急速に注目を集めた。国内小売業においても、需要予測・在庫最適化・価格最適化・顧客対応の自動化など広範な業務への適用が期待され、IT 投資額は年々増加している。リテールテック関連展示会においても生成 AI 活用サービスの出展が急増し、経営層の関心は高まり続けている。

しかしながら、国内外において生成 AI 投資の明確な ROI（投資対効果）が確認された事例は極めて少ない。MIT の報告では AI パイロットプロジェクトの 95%が有意な成果を出せず、RAND は AI 関連開発プロジェクトの 80%以上が失敗しているとする。S&P Global の調査でも PoC から本番導入において 46%のプロジェクトが中止に至っている。ガートナーのハイプサイクル（2025 年版）によれば、日本における生成 AI は「過度な期待のピーク期」にあり、LLM プラットフォームが主流として採用されるまでには 2～5 年を要すると分析されている。

こうした「生成 AI への高い期待」と「財務的成果が定量的に確認された事例が著しく少ない現実」の乖離こそが、本研究の出発点である。

研究目的と仮説

3つのアプローチ

本研究は以下の3つのアプローチを組み合わせた複合的な実証研究である。

1. **定量分析（財務）**：日本の上場企業全体を対象に、有価証券報告書のテキスト分析と財務業績データを組み合わせたラグ OLS 回帰分析
2. **定量分析（アンケート）**：小売業従事者を対象とした一次調査による現場実態の把握
3. **定性分析（文献・事例）**：NRF Big Show 2025・2026 の知見を含む海外先進事例との比較考察

研究仮説

表 1 研究仮説の一覧

仮説	内容
H1 (主仮説)	小売企業が生成 AI に多額投資しても、翌期の業績向上は限定的である
H2 (開示と業績)	AI・DX 関連キーワードを多く開示する企業は翌年の ROA・営業利益率が高い傾向がある
H3 (小売業特性)	小売業では AI・DX 開示が著しく少なく、キーワードと業績の統計的関係が検出されない
H4 (業種比較)	AI・DX 開示と業績の関係は IT 企業と小売業で方向・強度が異なる
H5 (現場調査)	小売業では関心は高いが実活用は限定的で、導入障壁が存在する

データと分析手法

データソースとスコア設計

一次データとして、金融庁 EDINET API v2 を通じて取得した日本の上場企業（9,547 件）の有価証券報告書全文（2022～2025 年度）を使用した。財務データは J-Quants API から取得した。

有価証券報告書のテキストから以下の 3 種類のキーワードスコアを算出した。

表 2 キーワードスコアの定義

スコア名	対象キーワード	重みづけ
生成 AI キーワード	生成 AI・ChatGPT・LLM・大規模言語モデル等	3（高）
AI 全般キーワード	AI・人工知能・機械学習・深層学習等	2（中）
DX 関連キーワード	DX・デジタルトランスフォーメーション・デジタル化等	1（標準）
合算スコア	上記 3 スコアの合算	—

分析手法：ラグ OLS 回帰と 140 仕様全探索

分析の中核は **1 期ラグ設計**（t 期のキーワードスコアが t+1 期の財務業績を予測するか）である。単一仕様での推定ではなく、目的変数 4 種（ROA・営業利益率・純利益率・売上高成長率）×キーワード変数 4 種×コントロール変数の組み合わせからなる **140 仕様を全探索** することで、特定の変数組み合わせへの恣意的な依存（p-hacking）のリスクを低減した。

$$\text{outcome}(t+1) = \beta_0 + \beta_1 \cdot \text{keyword}(t) + \beta_2 \cdot \log(\text{売上高})(t) + \beta_3 \cdot \text{outcome}(t) + \beta_4 \cdot D_{\text{小売}} + \varepsilon$$

主要な知見

知見① 全業種：DX 開示と翌期 ROA の正の関係

全業種 9,547 件を対象とした分析では、DX 関連キーワードスコア（DX 開示スコア）と翌期 ROA の間に有意な正の関係が確認された（ $\beta \approx 0.13 \sim 0.16$ 、 $p < 0.005$ ）。この効果はコントロール変数の仕様変更・業種不明企業の除外・逆因果検証後も安定して維持された。

この効果は Spence（1973）のシグナリング理論と整合的である。DX 関連キーワードを積極的に開示する企業は、投資家・取引先・求職者に「デジタル変革に取り組む企業」というシグナルを発信することで信頼を獲得し、資本コストの低下・優良契約獲得・優秀人材確保という間接的な経路で ROA が改善する可能性がある。ただし、本研究はこの機序（メカニズム）を直接検証したわけではなく、シグナリング理論による説明は一つの解釈として提示するものである。

知見② 情報通信業：AI 投資の短期コスト負担

情報通信業では、AI 全般キーワードスコア（AI 開示スコア）と翌期営業利益率の間に有意な負の関係が確認された（ $\beta \approx -0.25 \sim -0.33$ 、 $p < 0.05$ ）。逆因果検証後に係数がさらに強まったことから（ β ： $-0.114 \rightarrow -0.245$ ）、これは業績悪化企業が AI 言及を増やすという逆因果バイアスが係数を下押ししていたことを意味する。すなわち、業績水準に関わらず AI 投資に伴うコストが短期業績を圧迫するという投資コスト効果が生じている可能性が高いことが示唆される。ただし、これは観察データに基づく解釈であり、因果関係を最終的に確定するものではない。

知見③ 小売業：統計的関係の不在

小売業（東証 33 業種コード 6100、 $n=870$ ）においては、AI・DX キーワードスコアと翌期財務業績の間に統計的に有意な関係が全く検出されなかった（140 仕様すべてで $p > 0.29$ ）。

表 3 業種別の主要な分析結果

業種	主な結果	解釈
全業種合算	DX 関連キーワードスコア→ROA： $\beta \approx +0.13$ 、 $p < 0.005$	DX 開示のシグナリング効果
情報通信業	AI 全般キーワードスコア→営業利益率： $\beta \approx -0.25 \sim -0.33$	AI 投資の短期コスト先行
小売業	全 140 仕様で $p > 0.29$ （非有意）	キーワード開示量自体が不足

業種	主な結果	解釈
ガラス・土石 / 非鉄金属	正かつ有意な仕様あり	プロセス産業型の即効性

この非有意の主因は「効果がない」のではなく、小売業の生成 AI キーワードスコアが著しく低水準（情報通信業の 75 パーセンタイル値比で約 1/13）であり、説明変数として機能する開示量そのものが存在しないことにある。

知見④ 素材製造業：限定的な正効果

業種別分析（24 業種）の結果、AI/DX 言及と翌期業績向上の有意な正の相関はガラス・土石製品・非鉄金属という素材製造業 2 業種にのみ確認された（ $\beta \approx +0.35 \sim +1.23$ ）。プロセス最適化型の AI 活用が主流であり、投資コストの回収が短期間で実現しやすい業種特性が背景にある可能性がある。ただし、サンプルサイズが小さく（ $n=55 \sim 88$ ）、確認的知見ではなく示唆的な発見として扱う。

知見⑤ シンプソンのパラドックス的構造

全業種合算での正効果は、業種別分析では再現されない（24 業種中、正かつ有意な業種=0 件）。これはシンプソンのパラドックスと類似した構造である。機械・卸売・小売業といった大規模業種の点推定値が正方向に集中しており、業種構成バイアスが全体の相関を増幅させている。

図 1 分析結果の概念図

業種	結果	補足
全業種合算	正効果（ $\beta \approx +0.13$ ）	—
情報通信業	負効果 （ $\beta \approx -0.25 \sim -0.33$ ）	AI 投資コスト先行
小売業	無効果（ $p > 0.29$ ）	開示量不足
その他大多数	無効果	
素材製造業一部	正効果（限定的）	プロセス最適化効果

この結果は「DX 開示を増やせば ROA が上がる」という単純な因果解釈が誤りであり、業種特性を踏まえた分析が不可欠であることを示している。

アンケート調査からの知見：「開示不在 ≠ 活動不在」

（本節は第 5 章の定量分析による知見①～⑤とは別に、第 7 章のアンケート調査から得られた知見をまとめたものである。）

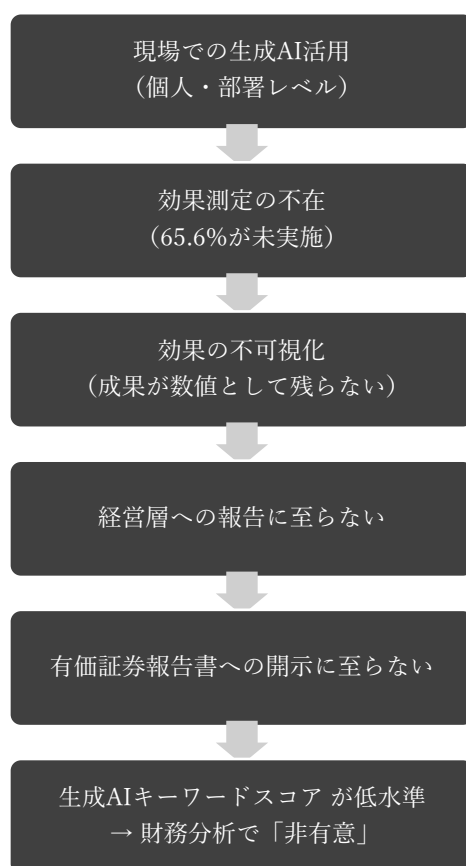
小売業従事者を対象としたアンケート調査（n=814）の結果、有価証券報告書上の開示水準の低さは「活動不在」を意味しないことが明らかになった。

表 4 アンケート調査の主要結果

項目	結果
生成 AI 効果測定の未実施率	65.6%（n=533）
測定実施群 vs 未測定群の評価差	全 7 観点で $\Delta = 0.59 \sim 0.82$ （ $p < .001$ ）
本社部門 vs 店舗部門の差（有意な観点）	業務生産性向上（ $p = .004$ ）・モチベーション向上（ $p = .032$ ）のみ
導入障壁 第 1 位	セキュリティ不安（21.5%）
導入障壁 第 2 位	成果の不可視性（17.9%）
導入障壁 第 3 位	教育・ルール整備の不十分さ（16.5%）

スクリーニング調査において小売業従事者 2,491 名のうち相当数が業務で生成 AI を何らかの形で活用していることが確認された。しかし活用者の 65.6%が効果測定を未実施であり、現場レベルの活用が定量的に把握・報告されない状況では、有価証券報告書への記載に至ることができない。

図 2 開示不在の構造的連鎖



※この連鎖は、財務分析（第 5 章）とアンケート調査（第 7 章）という別々のデータを組み合わせた解釈上の説明であり、単一の分析で直接検証された因果経路ではない。

これは「開示不在 ≠ 活動不在」という本研究の核心的な知見を示している。

アンケートでは、仮説 H5 をさらに検証可能な 6 つの命題（H5-1～H5-6）に分解して統計的に検証した。

表 4-2 H5-1～H5-6 の検証結果まとめ

仮説	内容	検証結果	主な根拠
H5-1	効果測定実施群は未測定群より Q8 の各評価が有意に高い	支持	全 7 観点で $p<.001$ ($\Delta=0.59\sim0.82$)
H5-2	Q8 の 7 観点間には強い正の相関がある	支持	全相関係数 $r=0.50\sim0.79$
H5-3	測定実施群と未測定群で評価理由 (Q9) の選択率に差がある	支持	8 項目全てで有意差 ($p<.05\sim<.001$)
H5-4	本社部門と店舗部門で Q8 評価に差がある	部分的に支持	7 観点中 2 観点のみ有意差 (業務生産性向上 $p=.004$ 、モチベーション向上 $p=.032$)
H5-5	効果測定の実施率が本社・店舗部門間で異なる	棄却	本社 33.8%・店舗 35.3%でほぼ同水準 ($p=.675$)
H5-6	評価理由 (Q9) の選択パターンが本社・店舗間で異なる	棄却	全 8 項目で有意差なし (全 $p>.10$)

効果測定の有無が評価を大きく左右する一方 (H5-1～H5-3)、部門差は個人軸の一部指標に限定され (H5-4)、組織的な意思決定は部門を問わず共通していた (H5-5・H5-6)。

考察

小売業での財務効果が検出されない構造的理由

有価証券報告書上の生成 AI キーワードスコアが低水準であることの原因は、①実際の生成 AI 実装水準が低いというシナリオと、②現場レベルでは試験的な活用が行われているが開示に至っていないというシナリオの 2 つが考えられた。アンケート調査の結果は②を支持しており、小売業では「個人レベルの活用（41.7%が主）→効果未測定→開示サイクルに乗らない」という連鎖が存在することが示された。

また、本研究のキーワード辞書に小売業固有の DX・AI 関連用語（オムニチャネル・POS 刷新・D2C・需要予測システム等）が含まれていないことも、測定バイアスの一因となっている可能性がある。

なお、この開示水準の低さは主仮説 H1（生成 AI 投資の効果は限定的）の解釈にも影響する。投資水準の代理指標となる開示自体にほとんど変動がないため、統計的関係の不在は「投資しても効果が限定的である」ことの証拠にはならない。H1 については本データのみでは支持・棄却いずれの判断も下せず、検証不能（判断保留）と結論づけるのが適切である。

NRF Big Show 2025・2026 との対比

世界最大の小売業カンファレンス NRF Big Show 2025・2026 では、欧米の大手小売企業による AI 活用の積極的な事例が多数報告された。しかしこの対比を「日本小売業の遅れ」と単純に解釈することには慎重を要する。NRF で紹介される事例は POS データ・顧客識別データ・オムニチャネル基盤が既に整備された欧米大手の話であり、日本の中堅小売業が置かれた出発点とは大きく異なる。

結論と実務的含意

仮説の検証結果まとめ

表 5 仮説の検証結果まとめ

仮説	検証結果	主な根拠
H1 (主仮説)	検証不能 (判断保留)	小売業では 140 仕様すべて $p>0.29$ だが、genAiScore が 全社でほぼゼロ＝投資水準の代理指標に変動がなく、 効果の有無を検証する前提条件が成立していない
H2 (開示と業績)	部分的に支持 (解釈に強い 留保が必要)	全業種合算では DX 関連キーワードスコア→ROA： β ≈ 0.13 で有意だが、24 業種中で正かつ有意な業種は 0 件。合算値は業種構成による合成効果である可能性が 高く、個別業種への一般化はできない
H3 (小売業特性)	支持	生成 AI キーワードスコア が情報通信業比約 1/13
H4 (業種比較)	支持	IT 業：負効果（有意）、小売業：検出不能（開示量不 足）
H5 (現場調査)	支持	効果測定未実施 65.6%・部門間実感差の確認

実務への示唆

小売業経営者へ： 生成 AI 活用の財務的効果を獲得するための最優先課題は**効果測定体制の構築**である。測定実施群は未測定群に比べ全 7 観点で有意に高い評価を示しており（ $\Delta = 0.59 \sim 0.82$ 、 $p<.001$ ）、「効果が見えない」という現場の声の多くは「測定していない」ことに起因する可能性が高い。

店舗部門への適用設計： テキスト生成・情報分析系 AI は本社業務（資料作成・経理・分析等）に直接的な恩恵をもたらしやすいが、接客・品出し・レジ等の物理的作業が中心の店舗業務では同じ施策が有効とは限らない。店舗向けには業務特性に合ったユースケース（需要予測・シフト最適化・異常検知等）を個別設計することが求められる。

本研究の限界と今後の課題

本研究には以下の限界が存在する。第一に、観察期間（2022～2025 年度）が短く、 $t+2$ 期以降の効果を捉えられていない。生成 AI の本格的な財務効果は、より長期の追跡研究が必要である。第二に、キーワード辞書が小売業固有の技術用語を含まない設計となっており、小

売業での測定バイアスが生じている可能性がある。第三に、ラグ回帰は真の因果関係の特定に限界があり、自然実験的なデザインや操作変数法の適用が今後の課題である。