

Appendix

A1. Data

Our dataset is available at <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5233143> and can also be requested from the co-authors.

Table A1. Journal Acronyms

AR	The Accounting Review	JAЕ	Journal of Accounting and Economics	JAR	Journal of Accounting Research
JF	Journal of Finance	JFE	Journal of Financial Economics	RFS	The Review of Financial Studies
ISR	Information Systems Research	MISQ	MIS Quarterly	JOC	Journal on Computing
AMJ	Academy of Management Journal	AMR	Academy of Management Review	SMJ	Strategic Management Journal
JIBS	Journal of International Business Studies	OS	Organization Science	ASQ	Administrative Science Quarterly
JM	Journal of Marketing	JMR	Journal of Marketing Research	MKS	Marketing Science
MS	Management Science	JCR	Journal of Consumer Research	JOM	Journal of Operations Management
MSOM	Manufacturing and Service Operations Management	OR	Operations Research	POM	Production and Operations Management

Criteria for Data Collection: A key decision in the data collection process is determining whether a paper qualifies as a machine learning measure-supported empirical study. First, this review considers only empirical studies that use secondary data or field experiments. We exclude studies based primarily on analytical modeling, simulation, and laboratory experiments. We also exclude the papers that focus on predictions but do not provide empirical insights. Second, we focus on papers that meaningfully incorporate machine learning. Specifically, we prioritize papers in which machine learning is central to the main analysis rather than used only for a miscellaneous post hoc analysis. Some papers include machine learning and empirical analysis as parallel modules, and we exclude these papers because neither module “supports” the other. Additionally, we screen out papers that employ heuristics/rules or directly adopt third-party datasets without implementing a machine learning procedure. We do so because the key parameters in heuristic methods are typically determined by researchers rather than learned from data, making them less amenable to machine learning control.

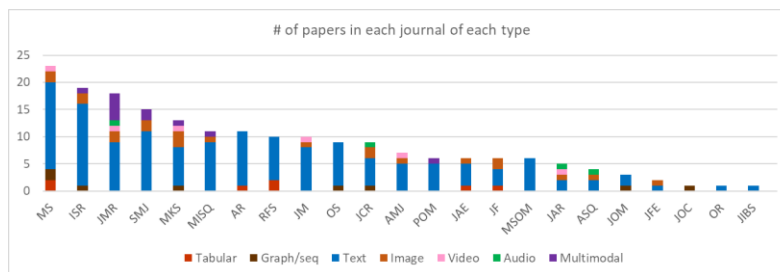


Figure A1. Paper Distribution by Journal and Data Type

A2. Cluster of Studies on Textual Measures

To visualize papers as textual data, we conduct document-level clustering on embeddings of the paper abstracts. Following Banker et al. (2024) and Reades and Williams (2023), we first preprocess the abstracts using the Natural Language Toolkit (NLTK) library (Bird et al. 2009), including lowercasing, removal of stop words, and lemmatization. We generate vector representations of each abstract (including the title) using the OpenAI model (text-embedding-3-large) to capture their semantic information. We apply K-Means clustering on the embeddings to group similar documents (Likas et al. 2003). To determine an appropriate number of clusters, we experiment with different values and evaluate the clustering quality using multiple metrics. The results in Figure A2 show that both the Silhouette Score (Rousseeuw 1987) and the Calinski–Harabasz Index (Calinski and Harabasz 1974) reach their optimal values when the number of clusters is three. Accordingly, we choose three clusters for our final model.

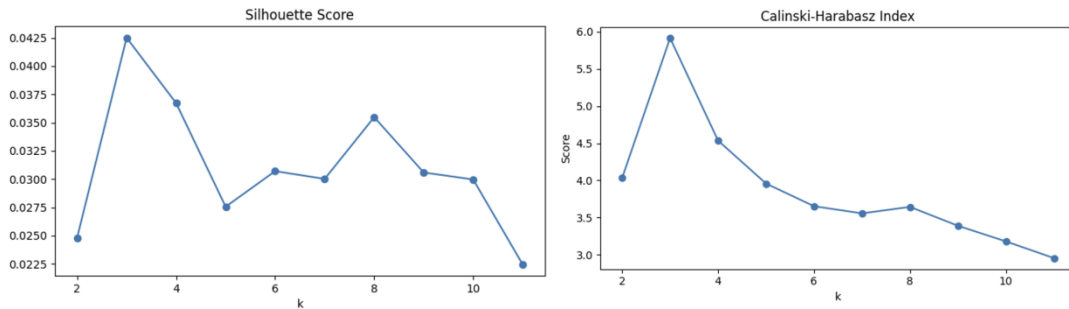


Figure A2. Evaluation of Clustering Results

To visualize the high-dimensional document embeddings, we first reduce the dimensionality to 50 using PCA, then further project them into 2D space using t-distributed Stochastic Neighbor Embedding (t-SNE) (Van der Maaten and Hinton 2008) and Uniform Manifold Approximation and Projection (UMAP) (McInnes et al. 2018), as reported in Figure 2. For each cluster, we also extract the most frequent and distinctive terms within each group, as reported in Table A2. We apply our best judgment to draw out the implications of these keywords.

Table A2. Top 10 Keywords of Clusters of All Textual Data Papers

Cluster	Keywords
Accounting/finance	market, return, financial, investor, information, disclosure, risk, analyst, ceo, firm
Marketing	service, patient, restaurant, quality, social, product, content, online, consumer, review
Management	learner, project, platform, firm, community, idea, distinctiveness, technology, cultural, product

A3. Roadmap for Machine Learning Measure-Supported Empirical Studies

Figure A3 integrates and summarizes our discussions and recommended practices in Sections 4.1 and 4.2 into a coherent roadmap. The roadmap first forks into Type I (machine learning to reduce human

coding) and Type II (machine learning for theoretical exploration) studies. While this decision depends largely on the specific research question, research context, and availability of data and labels, Figure A3 roughly differentiates the two types based on whether the study involves clearly defined and assessable constructs or whether objective labels are available.

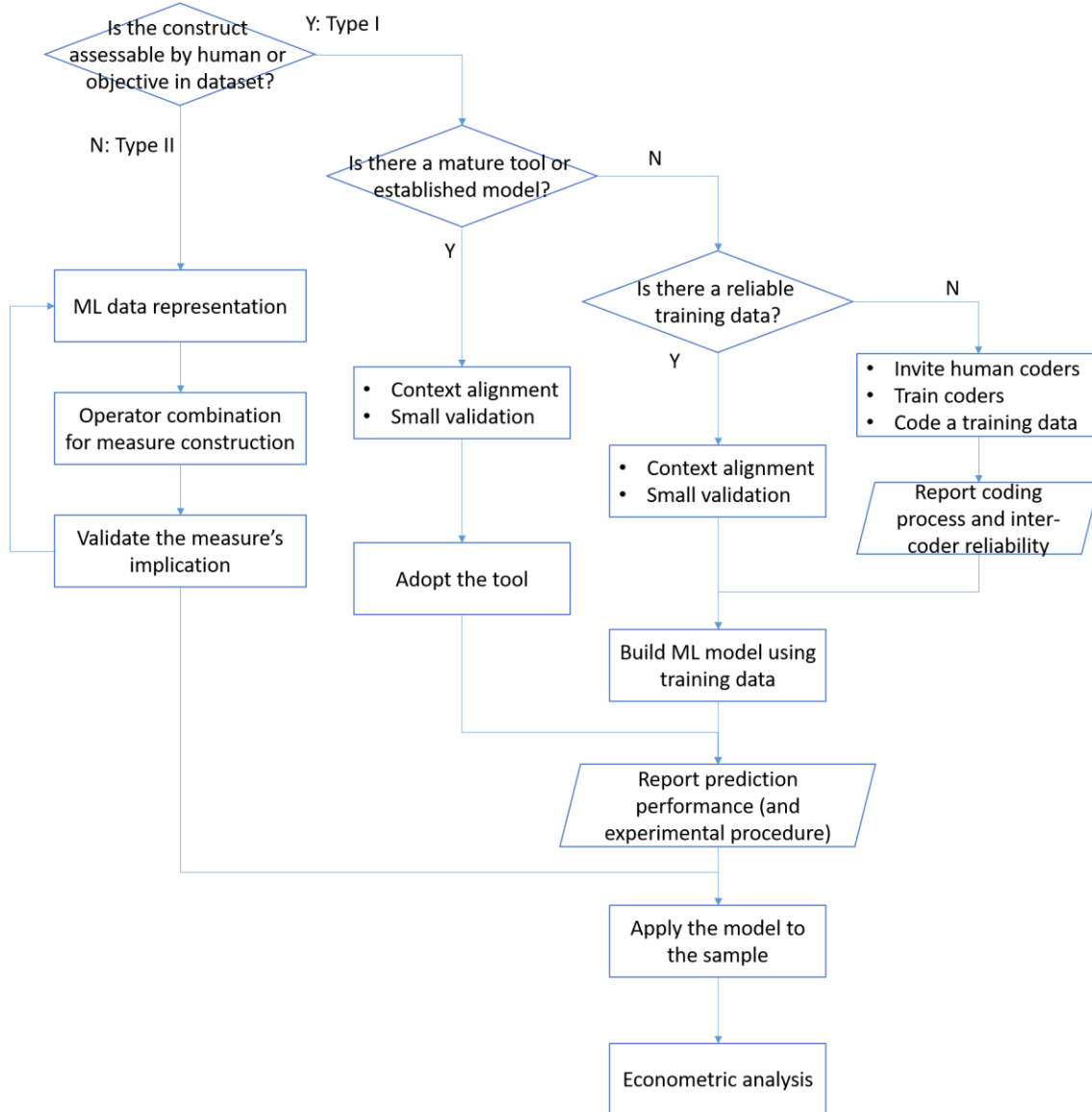


Figure A3. Roadmap for Machine Learning Measure-Supported Empirical Studies

For Type I studies, as discussed in Section 4.1.1, we distinguish three cases based on whether mature tools or training data are available. First, when there is no mature tool or no training data, researchers must create their own training labels in accordance with established coding guidelines (Bandalos 2018). Particular attention must be paid to coder qualifications, training procedures, and disagreement resolution (Puranam et al. 2021, Rhee 2024). Reporting thus often requires the number of coders, their training, the coding instructions, the inter-rater reliability (e.g., Cohen's kappa, Krippendorff's alpha) (Aggarwal and

Singh 2013), the number of annotators per item, and the method for resolving disagreements (e.g., majority vote, expert adjudication).

Second, when reliable annotated data exist (or when the prediction target is an objective value), researchers can adopt such datasets as training data. Then, the primary validity concern shifts to context alignment: whether the original annotation standards, data characteristics, and construct operationalization match the current research context. In these cases we suggest: 1) explicitly justifying why the existing dataset is appropriate for the current research context, addressing potential differences in domain or population; 2) conducting a small-scale validation study (e.g., 100-200 items) comparing the existing dataset's labels against human coding (Huang et al. 2018, Wang et al. 2022); and 3) discussing, if necessary, potential construct drift and its implications (Booker et al. 2023, Jiang et al. 2023).

Third, researchers sometimes use models (and parameters) developed in prior research or in commercial tools without retraining the models. As in the second case, the primary concern is still validity transfer: whether the adopted model adequately captures the construct in a new context. Unless the prediction target is objective or generic, we suggest: 1) justifying the context similarity between the adopted model and the applied problem; and 2) conducting a small validation analysis before adopting the model (Huang et al. 2018, Shin et al. 2020). It would also be necessary to document the software version.

While the aforementioned three streams of studies differ in how they obtain training data to build their machine learning models, once a model is established, they follow the same process. As discussed in Section 4.1.2, researchers need to report their model performance to ensure that the model captures the training signal. Although the specific data preprocessing steps and model-building details may not constitute the study's contribution, they may still need to be reported, including their data-splitting schemes and validation procedures (Song et al. 2019) and hyperparameters (Atalay et al. 2023), for the sake of replicability.

The Type II path in Figure A3 begins when researchers have no pre-existing human-coded training data and instead use machine learning to discover novel constructs from raw data. The measure emerges from an iterative process of representation learning, operator application, and conceptualization as discussed in Section 5.1. The central challenge is to justify the relevance of the derived measure to the intended construct. Since there is no human coding as a benchmark, researchers often need to show that the measure correlates with established proxies.

A distinctive feature of the Type II path is the iterative nature of construct development. As illustrated in Figure A3, the process loops back from construct validation to representation learning and operator design. However, papers rarely document these iterations explicitly.

The final two steps of the Figure A3 roadmap, "Apply the model to the sample" and "Econometric analysis," are common to both types of studies and reflect how machine learning measures are then fed

into econometric models to test causal or predictive relationships.

A4. Algorithm Customization

Table A3. Simplified Deep Learning Structures of Customized Machine Learning Models

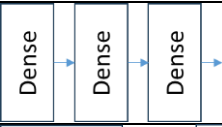
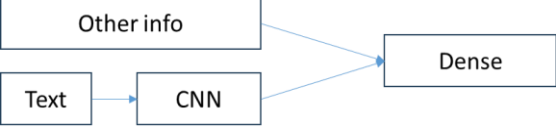
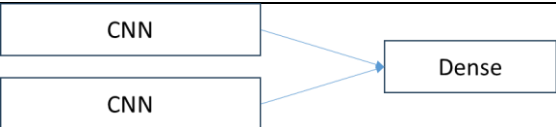
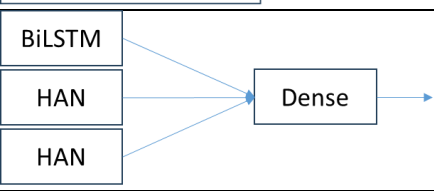
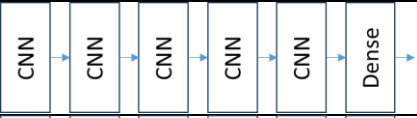
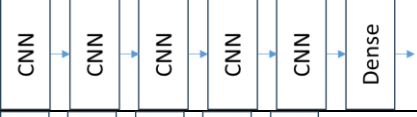
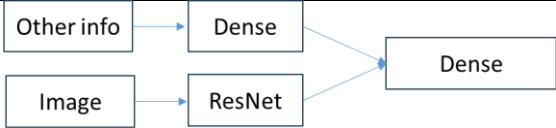
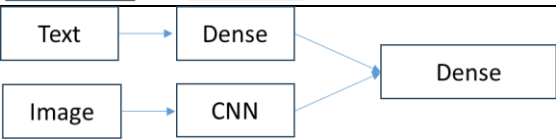
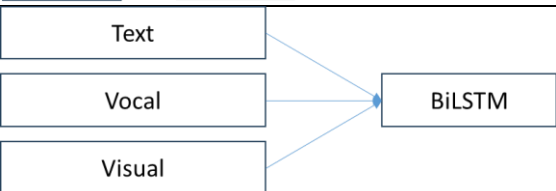
Paper	Data	Alg.	Model Structure	Type
(Kwark et al. 2021)	Text	DNN; KNN		Sequential
(Chakraborty et al. 2022)	Text	CNN+LSTM		Sequential
(Liu et al. 2019)	Text	CNN		Stack
(Liu et al. 2021)	Text	CNN		Stack
(Zhou et al. 2023)	Text	BiLSTM+HAN		Stack
(Zhang et al. 2022)	Image	CNN		Sequential
(L. Liu et al. 2020)	Image	CNN		Sequential
(Jiang et al. 2023)	Image	CNN		Sequential
(Aubry et al. 2023)	Image	CNN		Stack
(Hou et al. 2023)	Image	DNN		Stack
(Luo et al. 2024)	T+I+A	BiLSTM		Stack

Table A4. Time Complexity for Major Classic Machine Learning Algorithms

Algorithm	Training Complexity	Inference Complexity
Naïve Bayes	$O(n \cdot m)$	$O(m \cdot c)$
SVM	$O(n^2 \cdot m + n^3)$	$O(m \cdot n_{sv})$
Random Forest	$O(n_t \cdot n \cdot \log(n) \cdot m)$	$O(n_t \cdot d_t)$
Decision Tree	$O(n \cdot \log(n) \cdot m)$	$O(d_t)$
Logistic Regression	$O(ep \cdot n \cdot m)$	$O(m)$
Linear Regression	$O(n \cdot m^2 + m^3)$	$O(m)$

* n: number of samples; m: number of variables; ep: number of epochs; c: number of classes; d_t : tree depth; n_{sv} : number of support vectors; n_t : number of trees. Extracted from Chawla (2025).

Table A5. Time Complexity for Major Deep Learning Modules

Module	Inference Complexity per Layer
Convolutional Neural Network (CNN)	$O(k \cdot l \cdot m^2)$
Recurrent Neural Network (RNN)	$O(l \cdot m^2)$
Long Short-Term Memory (LSTM)	$O(l \cdot m^2)$
Self-Attention (Transformer)	$O(l^2 \cdot m)$
Restricted Self-Attention	$O(r \cdot l \cdot m)$

* l: sequence length; m: representation/model dimension; k: convolution kernel size; r: attention neighborhood size. Extracted from Vaswani et al. (2017). The time complexity for training deep learning models also needs to account for backward propagation, scaled by the sample size n and the number of epochs ep (Aibin 2025).

A5. Major Constructs

Table A6. Major Constructs in Using Machine Learning to Reduce Human Coding

Data Type	Training Source*	Construct	Exemplar Study
Tabular	Objective values	Firm earing	(Van Binsbergen et al. 2023)
Tabular	Objective values	Market haircut	(Jasova et al. 2024)
Graph	Objective values	Link	(Kang et al. 2023)
Image	Objective values	Face	(Chung et al. 2023)
Image	Objective values	Gender/race	(Anand and Kadiyalib 2024)
Image	Perceived by human	Facial expression	(Momtaz 2021)
Image	Perceived by human	Attractiveness	(Peng et al. 2020)
Image	Perceived by human	Trustworthiness	(Hsieh et al. 2020)
Image	Perceived by human	Image quality	(Li et al. 2023)
Image	Perceived by human	Sentiment	(Obaid and Pukthuanthong 2022)
Image	Objective values	Price	(Aubry et al. 2023)
Video/Multimodal	Objective values	Face	(Li and Xie 2020);
Video	Perceived by human	Facial expression	(Lin et al. 2021)
Video	Perceived by human	Pose	(Dávila and Guasch 2022)
Video	Perceived by human	Sentiment	(Van Den Oever and Shropshire 2024)
Multimodal	Perceived by human	Trustworthiness	(Luo et al. 2024)
Audio	Perceived by human	Vocal emotion	(Wang et al. 2021)
Audio	Perceived by human	Energy	(Yang et al. 2022)
Text	Perceived by human	Sentiment	(Huang et al. 2014);
Text	Perceived by human	Personality	(Liu et al. 2021)
Text	Perceived by human	Risk & uncertainty	(Chen et al. 2024)
Text	Perceived by human	Innovation	(Cong et al. 2024)
Text	Perceived by human	Regulation	(Kalmenvitz 2023)
Text	Perceived by human	Culture	(Li et al. 2021)
Text	Perceived by human	Organizational role	(Harrison et al. 2023)

Text	Perceived by human	Deception	(Hyde et al. 2024)
Text	Perceived by human	Content specificity	(Melumad et al. 2019)
Text	Perceived by human	Content style	(Ordenes et al. 2019)
Text	Perceived by human	Content quality	(Xu et al. 2021)
Text	Objective values	Medical	(X. Liu et al. 2020)

* In images or videos, human face, gender, and race may also need to be coded by humans, but we consider them relatively objective and easy to align with the intended construct to study.

Table A7. Major Constructs in Using Machine Learning for Conceptual Exploration

Data Type	Operator	Construct	Exemplar Study
Tabular	Splitting	Bidding strategy	(Adomavicius et al. 2012)
Graph	Splitting	Community	(Gagliardi et al. 2024)
Sequence	Splitting	Sequential pattern	(Packard et al. 2024)
Text	Splitting	Topic	(Archak et al. 2011)
Multimodal	Pooling	Voice amount	(Chang et al. 2023)
Text	Pooling	# Topics	(Hwang et al. 2019)
Text	Pooling	Narrativity	(Mukhopadhyay et al. 2022)
Image	Comparing	Style similarity	(Shi et al. 2023)
Video	Comparing	Engagement	(Yang et al. 2025)
Video	Comparing	Hemifacial asymmetry	(Banker, Ding, et al. 2024)
Audio	Comparing	Similarity	(Shi 2023)
Multimodal	Comparing	Image-text similarity	(Shin et al. 2020)
Text	Comparing	Agreement	(Chen et al. 2021)
Text	Comparing	Similarity	(Dyer et al. 2024)
Tabular	Comparing + Pooling (Intra-group)	Diversity	(Drake et al. 2024)
Graph	Comparing + Pooling (Intra-group)	Innovation decentralization	(Wu et al. 2019)
Text	Comparing + Pooling (Intra-group)	Diversity	(Choi et al. 2021)
Text	Comparing + Pooling (Intra-group)	Dispersion	(Huang et al. 2023)
Text	Comparing + Pooling (Intra-group)	Tightness	(Lee et al. 2022)
Image	Pooling + Comparing (1 vs. Group)	Uniqueness	(Orazi and Mannucci 2024);
Image	Pooling + Comparing (1 vs. Group)	Distinctiveness	(Banerjee et al. 2023)
Text	Pooling + Comparing (1 vs. Group)	Novelty	(Burtch et al. 2022)
Text	Pooling + Comparing (1 vs. Group)	Distinction	(Lee et al. 2023)
Text	Pooling + Comparing (1 vs. Group)	Differentiation	(Carlson 2023)

References

- Adomavicius G, Curley SP, Gupta A, Sanyal P (2012) Effect of information feedback on bidder behavior in continuous combinatorial auctions. *Manage. Sci.* 58(4):811–830.
- Aggarwal R, Singh H (2013) Differential influence of blogs across different stages of decision making: the case of venture capitalists. *MIS Q.* 37(4):1093–1112.
- Aibin M (2025) What is the time complexity for training a neural network using backpropagation? | baeldung on computer science. Retrieved (June 12, 2026), <https://www.baeldung.com/cs/backpropagation-time-complexity>.
- Anand P, Kadiyalib V (2024) Smoke and mirrors: impact of e-cigarette taxes on underage social media posting. *Mark. Sci.* 43(3):479–487.
- Archak N, Ghose A, Ipeirotis PG (2011) Deriving the pricing power of product features by mining consumer reviews. *Manage. Sci.* 57(8):1485–1509.
- Atalay AS, Kihal S El, Ellsaesser F (2023) Creating effective marketing messages through moderately

- surprising syntax. *J. Mark.* 87(5):755–775.
- Aubry M, Kräussl R, Manso G, Spaenjers C (2023) Biased auctioneers. *J. Finance* 78(2):795–833.
- Bandalos DL (2018) *Measurement Theory and Applications for the Social Sciences* (Guilford Publications).
- Banerjee M, Cole BM, Ingram P (2023) Distinctive from what? and for whom? deep learning-based product distinctiveness, social structure, and third-party certifications. *Acad. Manag. J.* 66(4):1016–1041.
- Banker RD, Ding H, Huang R, Li X (2024) Market reaction to CEOs' dynamic hemifacial asymmetry of expressions. *Manage. Sci.* 70(7):4847–4874.
- Banker RD, Huang R, Li X, Yan Y (2024) Strategy typology and cost structure: a textual analysis approach. *Prod. Oper. Manag.* 33(6):1285–1305.
- Van Binsbergen JH, Han X, Lopez-Lira A (2023) Man versus machine learning: the term structure of earnings expectations and conditional biases. *Rev. Financ. Stud.* 36(6):2361–2396.
- Bird S, Klein E, Loper E (2009) *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit* (O'Reilly Media, Cambridge).
- Booker A, Curtis A, Richardson VJ (2023) Investor disagreement, disclosure processing costs, and trading volume evidence from social media. *Account. Rev.* 98(1):109–137.
- Burtch G, He Q, Hong Y, Lee D (2022) How do peer awards motivate creative content? experimental evidence from reddit. *Manage. Sci.* 68(5):3488–3506.
- Caliński T, Harabasz J (1974) A dendrite method for cluster analysis. *Commun. Stat.* 3(1):1–27.
- Carlson NA (2023) Differentiation in microenterprises. *Strateg. Manag. J.* 44(5):1141–1167.
- Chakraborty I, Kim M, Sudhir K (2022) Attribute sentiment scoring with online text reviews: accounting for language structure and missing attributes. *J. Mark. Res.* 59(3):600–622.
- Chang HH, Mukherjee A, Chattopadhyay A (2023) More voices persuade: the attentional benefits of voice numerosity. *J. Mark. Res.* 60(4):687–706.
- Chawla A (2025) Time complexity of 10 ml algorithms. *Dly. Dose Data Sci.* Retrieved (May 11, 2026), <https://www.dailydoseofds.com/p/time-complexity-of-10-ml-algorithms/>.
- Chen D, Hu N, Liang P, Swink M (2024) Understanding the impact of trade policy effect uncertainty on firm-level innovation investment. *J. Oper. Manag.* 70(2):316–340.
- Chen K, Li X, Luo P, Zhao JLL (2021) News-induced dynamic networks for market signaling: understanding the impact of news on firm equity value. *Inf. Syst. Res.* 32(2):356–377.
- Choi J, Menon A, Tabakovic H (2021) Using machine learning to revisit the diversification–performance relationship. *Strateg. Manag. J.* 42(9):1632–1661.
- Chung JJ, Ding Y, Kalra A (2023) I really know you: how influencers can increase audience engagement by referencing their close social ties. *J. Consum. Res.* 50(4):683–703.
- Cong LW, Yang X, Zhang X (2024) Small and medium enterprises amidst the pandemic and reopening: digital edge and transformation. *Manage. Sci.*
- Dávila A, Guasch M (2022) Managers' body expansiveness, investor perceptions, and firm forecast errors and valuation. *J. Account. Res.* 60(2):517–563.
- Drake MS, Moon JR, Warren JD (2024) Classifying forecasts. *Account. Rev.*
- Dyer TA, Roulstone DT, Van Buskirk A (2024) Disclosure similarity and future stock return comovement. *Manage. Sci.* 70(7):4762–4780.
- Gagliardi L, Mariani M, Breschi S (2024) Temporal availability and women career progression: evidence from cross-time-zone acquisitions. *Organ. Sci.* 35(6):2178–2197.
- Harrison JS, Josefy MA, Kalm M, Krause R (2023) Using supervised machine learning to scale human-coded data: a method and dataset in the board leadership context. *Strateg. Manag. J.* 44(7):1780–1802.
- Hou JR, Zhang J, Zhang K (2023) Pictures that are worth a thousand donations: how emotions in project images drive the success of online charity fundraising campaigns? an image design perspective. *MIS Q.* 47(2):535–584.
- Hsieh TS, Kim JB, Wang RR, Wang Z (2020) Seeing is believing? executives' facial trustworthiness,

- auditor tenure, and audit fees. *J. Account. Econ.* 69(1).
- Huang AH, Leheavy R, Zang AY, Zheng R (2018) Analyst information discovery and interpretation roles: a topic modeling approach. *Manage. Sci.* 2833–2855.
- Huang AH, Zang AH, Zheng R (2014) Evidence on the information content of text in analyst reports. *Account. Rev.* 89(6):2151–2180.
- Huang H, Sunar N, Swaminathan JM, Roy R (2023) Do noisy customer reviews discourage platform sellers? empirical analysis of an online solar marketplace. *Manuf. Serv. Oper. Manag.* 25(6):2195–2215.
- Hwang EH, Singh PV, Argote L (2019) Jack of all, master of some: information network and innovation in crowdsourcing communities. *Inf. Syst. Res.* 30(2):389–410.
- Hyde SJ, Bachura E, Bundy J, Gretz RT, Sanders WG (2024) The tangled webs we weave: examining the effects of ceo deception on analyst recommendations. *Strateg. Manag. J.* 45(1):66–112.
- Jasova M, Laeven L, Mendicino C, Peydró JL, Supera D (2024) Systemic risk and monetary policy: the haircut gap channel of the lender of last resort. *Rev. Financ. Stud.* 37(7):2191–2243.
- Jiang J, Kelly B, Xiu D (2023) (Re-)imag(in)ing price trends. *J. Finance* 78(6):3193–3249.
- Kalmenovitz J (2023) Regulatory intensity and firm-specific exposure. *Rev. Financ. Stud.* 36(8):3311–3347.
- Kang M, Templeton G, Um SY (2023) Vulnerability diffusions in software product networks. *J. Oper. Manag.* 69(8):1342–1370.
- Kwark Y, Lee GM, Pavlou PA, Qiu L (2021) On the spillover effects of online product reviews on purchases: evidence from clickstream data. *Inf. Syst. Res.* 32(3):895–913.
- Lee JJ, Jung HJ, Park H (2023) Rare is beautiful? rareness, technology value, and the moderating role of search domain and knowledge maturity. *Prod. Oper. Manag.* 32(4):1019–1040.
- Lee JK, Junque De Fortuny E, De Fortuny EJ (2022) Influencer-generated reference groups. *J. Consum. Res.* 49(1):25–45.
- Li H, Simchi-Levi D, Wu MX, Zhu W (2023) Estimating and exploiting the impact of photo layout: a structural approach. *Manage. Sci.* 69(9):5209–5233.
- Li K, Mai F, Shen R, Yan X (2021) Measuring corporate culture using machine learning. *Rev. Financ. Stud.* 34(7):3265–3315.
- Li Y, Xie Y (2020) Is a picture worth a thousand words? an empirical study of image content and social media engagement. *J. Mark. Res.* 57(1):1–19.
- Likas A, Vlassis N, Verbeek JJ (2003) The global k-means clustering algorithm. *Pattern Recognit.* 36(2):451–461.
- Lin Y, Yao D, Chen X (2021) Happiness begets money: emotion and engagement in live streaming. *J. Mark. Res.* 58(3):417–438.
- Liu AX, Li Y, Xu SX (2021) Assessing the unacquainted: inferred reviewer personality and review helpfulness. *MIS Q.* 45(3):1113–1148.
- Liu L, Dzyabura D, Mizik N (2020) Visual listening in: extracting brand image portrayed on social media. *Mark. Sci.* 39(4):669–686.
- Liu X, Lee D, Srinivasan K (2019) Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning. *J. Mark. Res.* 56(6):918–943.
- Liu X, Zhang B, Susarla A, Padman R (2020) Go to youtube and call me in the morning: use of social media for chronic conditions. *MIS Q.* 44(1):257–283.
- Luo X, Jia N, Ouyang E, Fang Z (2024) Introducing machine-learning-based data fusion methods for analyzing multimodal data: an application of measuring trustworthiness of microenterprises. *Strateg. Manag. J.* 45(8):1597–1629.
- Van der Maaten L, Hinton G (2008) Visualizing data using t-sne. *J. Mach. Learn. Res.* 9(11).
- McInnes L, Healy J, Saul N, Großberger L (2018) UMAP: uniform manifold approximation and projection. *J. Open Source Softw.* 3(29):861.
- Melumad S, Inman JJ, Pham MT (2019) Selectively emotional: how smartphone use changes user-generated content. *J. Mark. Res.* 56(2):259–275.

- Momtaz PP (2021) CEO emotions and firm valuation in initial coin offerings: an artificial emotional intelligence approach. *Strateg. Manag. J.* 42(3):558–578.
- Mukhopadhyay S, Kumar V, Sharma A, Chung TS (2022) Impact of review narrativity on sales in a competitive environment. *Prod. Oper. Manag.* 31(6):2538–2556.
- Obaid K, Pukthuanthong K (2022) A picture is worth a thousand words: measuring investor sentiment by combining machine learning and photos from news. *J. Financ. Econ.* 144(1):273–297.
- Van Den Oever K, Shropshire C (2024) More than a feeling: how board member displays of anger and happiness influence strategic decisions. *Acad. Manag. J.* 67(6):1525–1552.
- Orazi DC, Mannucci PV (2024) Disentangling audiences' reactions to creative content and creative packaging. *Strateg. Manag. J.* 45(13):2763–2790.
- Ordenes FV, Grewal D, Ludwig S, Ruyter K De, Mahr D, Wetzels M (2019) Cutting through content clutter: how speech and image acts drive consumer sharing of social media brand messages. *J. Consum. Res.* 45(5):988–1012.
- Packard G, Li Y, Berger J (2024) When language matters. *J. Consum. Res.* 51:634–654.
- Peng L, Cui G, Chung Y, Zheng W (2020) The faces of success: beauty and ugliness premiums in e-commerce platforms. *J. Mark.* 84(4):67–85.
- Puranam D, Kadiyali V, Narayan V (2021) The impact of increase in minimum wages on consumer perceptions of service: a transformer model of online restaurant reviews. *Mark. Sci.* 40(5):985–1004.
- Reades J, Williams J (2023) Clustering and visualising documents using word embeddings. *Program. Hist.*
- Rhee L (2024) CEO attentional vigilance: behavioral implications for the pursuit of exploration. *Acad. Manag. J.* 67(6):1463–1487.
- Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20(C):53–65.
- Shi Y (2023) A change of tune: the democratization of market mediation and crossover production in the u.s. commercial music industry. *Adm. Sci. Q.* 68(2):319–354.
- Shi Z, Liu X, Lee D, Srinivasan K (2023) How do fast-fashion copycats affect the popularity of premium brands? evidence from social media. *J. Mark. Res.* 60(6):1027–1051.
- Shin D, He S, Lee GM, Whinston AB, Cetintas S, Lee KC (2020) Enhancing social media analysis with visual data analytics: a deep learning approach. *MIS Q.* 44(4):1459–1492.
- Song T, Huang J, Tan Y, Yu Y (2019) Using user- and marketer-generated content for box office revenue prediction: differences between microblogging and third-party platforms. *Inf. Syst. Res.* 30(1):191–203.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. *Adv. Neural Inf. Process. Syst.* 5999–6009.
- Wang X, Lu S, Li X, Khamitov M, Bendle N (2021) Audio mining: the role of vocal tone in persuasion. *J. Consum. Res.* 48(2):189–211.
- Wang Y, Qin MS, Luo X, Kou Y (2022) Frontiers: how support for black lives matter impacts consumer responses on social media. *Mark. Sci.* 41(6):1029–1044.
- Wu L, Lou B, Hitt L (2019) Data analytics supports decentralized innovation. *Manage. Sci.* 65(10):4863–4877.
- Xu Y, Armony M, Ghose A (2021) The interplay between online reviews and physician demand: an empirical investigation. *Manage. Sci.* 67(12):7344–7361.
- Yang J, Xie Y, Krishnamurthi L, Papatla P (2022) High-energy ad content: a large-scale investigation of tv commercials. *J. Mark. Res.* 59(4):840–859.
- Yang J, Zhang J, Zhang Y (2025) Engagement that sells: influencer video advertising on tiktok. *Mark. Sci.* 44(2):247–267.
- Zhang S, Lee D, Singh PV, Srinivasan K (2022) What makes a good image? airbnb demand analytics leveraging interpretable image features. *Manage. Sci.* 68(8):5644–5666.
- Zhou J, Zhang Q, Zhou S, Li X, Zhang X (2023) Unintended emotional effects of online health communities: a text mining-supported empirical study. *MIS Q.* 47(1):195–226.