Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD29153

The Landscape of Modern AI in Finance: A Review of Techniques, Risks, and Governance for Scalable Deployment

Hanlin Zhou[1], Junyi Huang[2*]

[1]Shenzhen College of International Education, Shenzhen, China
[2]University College London, London, United Kingdom
**Corresponding Author. Email: junyi.huang.24@ucl.ac.uk*

Abstract. This article presents financial AI as 'method-application' mapping and minimal evaluation toolkit. From case-backed evidence, we find: calibrated ensembles remain the baseline; time-sequence and graph model prioritize when signal iss nonlinear or relational; RL is feasible solely in conservative, offline policy and credible emulators; LLMs superiority focus on searching, boundaries and auditability. Cross-scenario risks include data availability and privacy, label scarcity and drifting, OOD weakness and miscalibration. Mitigations focus on MRM, fairness, lineage and versioning, monitoring, and cost-latency trade-offs. Ultimately, we conclude a deployment checklist: data contracts; interpretable, monotonic or constrained models as needed; robustness and calibrated training; causal and counterfactual evaluation; and institution-level MLOps, LLMOps. Hence, empowering system evolution from 'usable' to 'trustworthy' at scale.

Keywords: Financial AI, Machine Learning, Fintech, Risk Management, Trustworthiness

1. Introduction

The financial industry is currently being reshaped by advances in computing, data availability and artificial intelligence (AI). Both retail finance and wholesale business require institutions to process heterogeneous, vast volumes and high-speed data under strict regulatory and time constraints – including structured time-series data (e.g., open-high-low-close-volume (OHLCV) quotes and limit order books (LOB) messages) [1], as well as the unstructured text (e.g., corporate disclosures and regulatory documents). Conventional rule-driven and linear statistical methods demonstrate mediocrity with rising customer expectations on personalization and 'always-online' services when coping with market instability, complex interactions, and confrontational behavior (e.g., fraud) [2,3].

Modern AI, including supervised and unsupervised learning, deep representation learning, reinforcement learning (RL), and Natural Language Processing (NLP), offers better capacity in pattern discovery, sequential decision-making, and language understanding: discovering risk-profit-driven factors, identifying anomaly flow, supporting conversational services, and extracting knowledge to assist front-office business and control.

This review focuses on the demonstrated or near-term feasible AI approaches in financial scenarios, systematically mapping them into four critical business fields: (1) investment

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.LD29153

management and algorithmic trading; (2) risk management; (3) customer service and marketing analytics; and (4) regulatory technology and compliance.

The goals of this article are to (i) present a clear mapping of 'Method-Application' over data modalities; (ii) summarize evidence of value as well as the typical failure modes. (iii) Clarify critical challenges related to trustworthiness, fairness, and governance. (iv) Identify research gaps, discussing causal inference, privacy-preserving learning and the potential contribution in optimizing stability by retrieval-augmented-generation (RAG).

2. Evolution of AI and technology basis

2.1. History review

This section recalls the previous applications of AI in financial scenarios. The evolution went as: (1) Expert Systems (2) Statistic and ensemble learning (3) deep and representation learning (4) RL for sequential decision-making (5) pretraining and LLMs.

(1) Rule-based and 'Expert System'

Earlier applications originated from knowledge engineering: constructed an explainable reasoning base and decision-making trees by encoding the decision logic of those experts in the fields and the regulatory rules. The result has been deployed in credit underwriting, anti-money laundering (AML) & sanctions screening and checklist-based operational risk controls. These systems are explainable and audit-friendly. Nevertheless, lack of adaptability to unstable market, complex interactions and novel confrontations (rapidly evolving fraud patterns) leads cost to soar exponentially as the business complexity increases [4].

(2) Statistical Learning and Ensemble Learning

As structured data on accounts, transaction, financial statements expanded, the financial industry widely adopted methods including logistic regression, discriminant analysis, support vector machines (SVMs) and tree ensembles (Random Forest (RF) and gradient boosting decision trees (GBDT)). These supports at credit scoring, default prediction, fraud detection and client segmentation. This era emphasizes robust feature engineering and statistical in-/out-of-sample validation.

(3) Deep and Representation Learning

Since the 2010s, enhancement in computation resources and data scale has stimulated the adoption of deep learning in industry. (i) Time-series modelling, convolutional neural network (CNN), recurrent neural network (RNN), long short-term memory (LSTM), gated recurrent unit (GRU), as well as the following attention mechanism and state-space models (SSM) were applied in non-linear modelling for price and flow dynamics volatility, Value-at-Risk (VaR) prediction, anomaly detection. (ii) With regard to unstructured data, word embeddings, contextualized language models and end-to-end text extraction for sentiment was applied to annual reports, announcement, research notes, customer-service dialogues emotion, event extraction and knowledge-graph construction. (iii) Multimodal fusion—co-modelling by structured indicators and text evidence began to applied in due diligence and reputation risk control.

The introduction of deep learning significantly increases the representational capability of high dimensional, non-linear and weakly structured signal, however, sacrificing the interpretability, bringing distribution drift and increasing data governance expense.

(4) Reinforcement Learning and Sequential Decision-making

To cope with dynamic environment and feedback delay tasks including trading, market-making, portfolio rebalancing, RL has been combined with environment emulation to optimize trading path

DOI: 10.54254/2755-2721/2025.LD29153

and inventory risk; In terms of fraud and policy, policy-gradient and offline RL were adopted in attempts to balance cost and profit. Nonetheless, RL faces challenges in offline data basis, controllability and off-policy evaluation (OPE), demanding human-machine cooperation and conservative strategy sets.

(5) Foundation-model Era

Revolutionized attention-mechanism-based Transformer model and pretraining brought great changes to the NLP area and rapidly became the industry standard, extended into multimodal models including code, time-series and etc. In the financial context, domain-adaptive pretraining and RAG is applied to power compliance Q&A, report generation and information extraction; Instruction-tuned conversational-models optimize efficiency and consistency on client-end.

2.2. Core AI technology illustration

This section takes the most general, universal technology elements and evaluation criteria in financial scenarios to construct a lexicon for the following chapters.

2.2.1. Supervised learning (classification/regression)

Supervised learning is general in financial risk control and marketing dealing with structured accounts, transactions and customer portraits. Generalized linear model (GLM), tree ensembles (RF etc.) and multilayer perceptron (MLP) are widely applied in credit scoring, default prediction, fraud detection, limits and risk exposure regression. Compared to traditional single tree decision, RF demonstrates superiority in precision and robustness [5]. Hence, GBDT such as extreme gradient boosting (XGBoost) and light gradient boost machine (LightGBM) can efficiently achieve balance between time and accuracy in large scale, feature scarce scenarios [6,7].

2.2.2. Unsupervised learning and anomaly detection

Unsupervised methods are mainly deployed in unlabeled transaction flow, interaction logs and networked data to complete customer segmentation, community & group discovery and suspicious pattern identification. Typical models include clustering and density method; One-class SVM [8], isolation forest [9] and representation-learning-based autoencoder (AE) [10]; Relational data coped with graph-anomaly detection to utilize structured information [11].

2.2.3. Semi-supervised and self-supervised learning

Semi and self-supervised learning have been broadly adopted to address label scarcity and expensiveness, vast unlabeled data. Typical techniques include label and consistency regularization [12], FixMatch [13], contrastive learning (e.g., Simple Framework for Contrastive Learning of Visual Representations (SimCLR)) [14], masked modeling and sequence-oriented data.

Notice the pseudo-label-noise enlargement and class-imbalance effects need to be controlled [15]; Under regulatory scenarios, these methods are generally combined with federated learning (FL) and privacy-preserving computing, expanding data coverage and enhancing generalization performance without violating data-sharing constraints [16,17].

DOI: 10.54254/2755-2721/2025.LD29153

2.2.4. Time-series modelling (sequence and events)

Financial time sequence includes price and trades, limit-order-book messages and account timelines, are mainly utilized for multi-step predictions, fluctuations & risks measurement, anomaly detection and change-point discovery. Temporal convolutional network (TCN) and LSTM, GRU are often used as baseline models [18-20]. The latest solution is to adopts attention-based sequence models (e.g., Temporal Fusion Transformer (TFT)) and SSM to capture long-range dependence and irregular sampling [21,22]. When linear and non-linear dynamics co-exist, autoregressive integrated moving average (ARIMA) and generalized autoregressive conditional heteroskedasticity (GARCH) are deployed with neural networks [23-25].

2.2.5. Graph learning

Financial interaction data naturally form a heterogeneous relation network, connecting accounts, vendors, devices and supply chain entities. Graph learning is applied in AML, group fraud identification and risk propagation analytics. Representative architectures include graph convolutional network (GCN), graph-sample-and-aggregate (GraphSAGE) and graph attention network (GAT) [26-28]; Relation and heterogeneous graph convolutional networks (GNNs) are used in multi-category nodes and relation scenarios [29]; In label scarcity scenario, graph contrastive learning outperforms self-supervised learning [30]. Lightweight embeddings such as Node2Vec and Deepwalk are operable feature generation methods for downstream categorizers [31,32]. Moreover, GCN variants and application research have been explored in cross-domain such as recommender systems [33].

2.2.6. NLP/LLM and retrieval

Financial text workloads cover various tasks: public announcement, annual report, customer service analytics etc. Domain-adaptive pretraining and instruction tuning (normally assists with parameter-efficient fine-tuning (PEFT) and low-rank adaptation (LoRA)) are applied to align basic models to industry terminology and tasks [34-36]; To cope with knowledge-dense tasks, RAG pipeline is used to establish citations for corporate language sources, executing 'searching-re-ranking-generation' [37-39].

2.2.7. Reinforcement learning and constrained optimization (overview)

Sequential decision problems are widely present in trade execution, market-making, inventory & capital management and impede strategy making. They are indicated as partially observable, delayed feedback along with risk, budget constraint.

Techniques include policy gradient, actor-critics methods are applied in online control; offline RL is applied to logged data [40,41]. Batch-constrained Q-learning (BCQ) and conservative Q-learning (CQL) are the classic approaches to mitigate extrapolation error [41,42]. RL with constraint or safe RL controls tail risk via conditional VaR (CVaR) or explicit constraint [43].

DOI: 10.54254/2755-2721/2025.LD29153

3. Applications by business domain

3.1. Investment and algorithmic trading

3.1.1. Data and tasks

Prime data includes OHLCV and LOB microstructure, fundamentals and financial indicators, public opinions and press texts, other data (macro high frequency, satellite, geolocation and app usage). Core tasks cover: (1) Alpha discovery and cross-sectional selection; (2) prices, volatility, and dependence predictions; (3) Portfolio construction and risk distribution (mean-variance, square difference, risk parity, factor exposure constraint, expected shortfall (ES) and conditional VaR constraints); (4) Execution Optimization and market making (Minimize impact cost, inventory, price-difference management); (5) Strategy assessment and storage management.

3.1.2. Methods

Signals from linear factors and ensembles to deep sequence and hybrid models; Adopt 'text-to-feature' pipeline for emotion and events extraction [44]. Adopt robustness and cost-aware portfolio optimization while introducing level-dependence and scenario generation simultaneously [45]. Deploy policy gradient and offline RL to achieve self-adaptive execution; Co-controlling stock and price difference in emulation environment [46,47].

3.1.3. Evaluation and governance

Avoid data snooping by taking walk-forward or rolling-origin back-testing. Critical indicators including information coefficient, information rate, hit-rate, turnover/capacity, cost/market impact, Sharpe/Sortino/Calmar. Utilize the deflated and probabilistic Sharpe ratios [48, 49] to control multiple testing [50]. Post-hoc attribution, probability of back-test overfitting and reality check [51,52].

3.1.4. General failure and mitigation

Distribution drift caused by market shifting, margin decline caused by trade congestion, time-effectiveness decline in text emotion and wrong report. Mitigation initiatives include small sample steady estimation, over-market/period validation, regularization alpha lineage management, hedging and dimensionality and human-machine cooperation downline standard.

3.2. Risk management (credit/market/maneuver/mobility)

3.2.1. Credit risks

Credit risk scenarios cover probability of default (PD), loss given default (LGD), exposure at default (EAD), credit value (CV) estimation, default warning, pricing and limit setting, and early harvesting [53]. In terms of engineering, scorecard and generalized linear models run parallel with gradient boost tree (e.g., LightGBM) and multilayer perceptron[7]; Relational network features and semi/self-supervised signal are introduced for thin-file small and medium-sized enterprises (SMEs) and long-tail segments, while limit-increases adopt uplift modeling [54].

DOI: 10.54254/2755-2721/2025.LD29153

3.2.2. Market risks

Tasks: Measure VaR and ES, analyze tail-dependence data, operating scenarios analysis, stress test, and build x-value adjustments (XVA) approximation [55].

Method: Adopt statistical-Machine Learning fusion (e.g., GARCH, Long memory and Deep network), quantile regression and modelling [24]. Relevant structure can be constructed by copula and graph method; Generative and emulative technology applies in scarce scenario generation.

Evaluation: Execute VaR Back-testing (default counting, 'traffic light'), report Quasi-Likelihood loss (QLike) and Pinball loss; Model risk management (Model variation, challenges and independent validation).

3.2.3. Maneuver and fraud risks

Tasks: Trade and transaction fraud, account takeover, device biometric identification anomaly, internal-misuse deviations.

Method: Supervised two classifications with anomaly detection, graph learning (loop, group detection), FL and multi-institution cooperation [3,9]; Strategy assignment integrates rules and models, assist with real-time interdiction and human inspection.

Indicators: Under polarized imbalance, focuses on precision-recall AUC (PR-AUC), recall@FPR, cost-sensitivity indication (per impede return, missing cost) [56].

3.2.4. Mobility risks and money dispatching

Tasks: Daily and overnight position prediction under liquidity coverage ratio (LCR), net stable funding ratio (NSFR) constraint and buffering, deposit optimization.

Method: Time-series prediction and constraint optimization; Deploy offline RL with Market-making, stock as strategy backup and set safety thresholds [41].

3.3. Customer services and marketing analysis

3.3.1. Task pattern

Customer categorization and portrait, Churn prediction, customer lifetime value (CLV) estimation, Next-Best-Action, offer recommendation, AML and customer risks distribution in Know Your Customer (KYC), customer services automation.

3.3.2. Method

Supervised/Semi-Supervised: GBDT/Deep with transition/Self-supervised to optimize cold start; Apply uplift model to distinguish effective customers [54].

Recommendation and research: recall-coarse-rank-fine-rank, enhancing target balancing via knowledge distillation and multitasking learning [57,58].

3.3.3. Evaluation

Marketing: Area-under-the-uplift-curve (AUUC), Qini coefficient (Qini), ROI increase, long-term holding and rebuy; Multi-Armed Bandit (MAB) [59].

DOI: 10.54254/2755-2721/2025.LD29153

Customer Services: First-contact resolution (FCR), average handle time (AHT), containment, customer satisfaction (CSAT)/ net promoter score (NPS); Generative faithfulness, trackable evaluation and risk tagging.

Regulatory and Privacy: Personally identifiable information (PII) de-identification, agreement and intention limitation, minor doable feature set; Fair evaluation to individual and group, as well as embedding interpretable evidences (SHAP, counterfactuals) into customer communications [60,61].

3.4. Regulatory technology and compliance

3.4.1. Scenarios and defective

KYC, AML watchlist and transaction monitoring, suspicious activity report (SAR) drafting and case management, requiring low miss and false-alarm rates [62]. Track and follow the regulation variation, completing the 'obligation-to-control-to-evidence' mapping [63]; Execute model risk management: inventory, change and version control, independent validation unit, continuous monitoring (including conversational safety in LLMs).

3.4.2. Technology and process

Information extraction including entity, relation, event extraction, obligation terms analysis as well as the RAG with search re-ranking supports auditable answers with citations [37]; Graph analytics to output subgraph evidence [11]; Machine Learning with rules to assure compliance and recall, human review near thresholds.

3.4.3. Governance and measurement

Lineage and audit trails; Prompt and knowledge versioning; Quality and risk indicators: missed and incorrect reports, time-to-resolution, closure rate; Policy consistency check and 'red line' answer-rejection [64]. Deployment and running: Identity and access management (IAM), data residency and overseas transmission control; Canary releases and rollback [65].

4. Challenges & limitations

This chapter sorts out the possible critical bottlenecks for AI to be utilized in financial scenarios. The chapter is constructed by four sections including data, models, governance and engineering.

4.1. Data and data governance

Availability and Compliance: Constraints in overseas and inter-institutional data sharing; personally identifiable information (PII). Label scarcity and latency: Contract breaching, fraud cases, regulatory violation are rare and inspection-delayed; Easy to trigger purpose leakage and evaluation bias [66]. Sample and representativeness bias: Outdated strategy and rules leads to selection bias; Long-tail customer groups and cold start issue [67]. Imbalance and drift: Macro and market shifting, product strategy changing leads to covariate and concept drift [2]. Data quality and spectrum: Lack, anomalies, type difference; lack of data contracts and trackable lineage [65,68].

DOI: 10.54254/2755-2721/2025.LD29153

4.2. Model and method

Interpretability and Auditability: The 'black-box' characteristic makes deep learning models lack interpretability and auditability. Rules and policy may be inconsistent due to the inability to interpret model behaviors [69]. Steadiness and generalization: Sample demonstrates weakness to out-of-distribution (OOD); Adversarial examples and data poisoning impacts on the overall model steadiness and robustness [70,71]. Small sample over-fitting and data snooping. Calibration and uncertainty: Overconfident in confidence range, threshold-selection error [72]. Sequence and consequences: Information leakage and lookback bias; Strategy and data interact, leading to consequences chaos [73].

Specifically for RL, Offline-data bias, OPE has limitations [74]. With regard to LLMs, hallucinations and jailbreaks are weakness and may threaten the model performances, triggering further issues [75,76].

4.3. Organization and governance

Responsible-accountable-consulted-informed (RACI) ambiguity causes decision errors. Institutionalize MRM, balancing fairness evaluation between individuals and groups [77]; Data residency, consent and purpose limitations system building. Lineage and audit trails construction and LLM output appeal processes are demanding [78].

4.4. Engineering

Time-effectiveness and availability: low-latency service-level agreement (SLA) and feature time-consistency [79,80]; Versioning and reproducing: Data, feature, model, prompt and knowledge versioning [63,79,80]; Monitoring and alerting: full-stack monitoring at 'input-feature-prediction-business KPI' [78,81]; Canary release, rollback and kill switches [82]; Inference cost and freshness trade-offs [77,81].

5. Expectations and future research

5.1. Data and privacy

Privacy-preserving RL and cooperative paradigms: FL, differential privacy (DP), secure multi-party computation (MPC), homomorphic encryption (HE) practical deployment in inter-institutional cooperation [16,17]; Mitigate label scarcity by combining federated searching and self-supervised or weak-supervised learning.

Data contract and trackability: Standard locking among data contract, Schema calibration, SLA and data quality (DQ) gates; Record global chain lineage and variation to support audit and traceback [83].

Steady sampling and data crafting: multi-level sampling and weighting to mitigate selection bias [84]; Utilize generative crafted-data within boundaries for rare scenario test and steadiness test.

DOI: 10.54254/2755-2721/2025.LD29153

5.2. Model methodology

5.2.1. Interpretability and constraint learning

Prioritize 'interpretable models and monotonic, business-logic constraints in regulatory-sensitive domains. Construct decision-evidence pack by SHAP, Integrated Gradients and counterfactual explanations [58].

5.2.2. Steadiness and calibration

Combining Distributionally Robust Optimization (DRO) and Invariant Risk Minimization (IRM) with adversarial training to enhance robustness under distortion and distribution drifting [85]. To make the model probabilistic outputs explainable, probability calibration including temperature scaling and isotonic regression should be done [70]. Simultaneously, OOD detection and warning mechanism should be constructed to identify irregular input [70], triggering downgraded or human inspection. Hence, forming closed loop of 'steadiness-calibration-monitoring'.

5.2.3. Consequences and evaluations

Deploy double Machine Learning (DML), causal forest and instrumental variables (IV) to mitigate error chaos [88]. Hence, getting estimation for consequences. Meanwhile, introduce counterfactual back-testing examines whether the strategy truly add value.

5.2.4. 'Conservative and constraint'-standard for RL

With regard to tasks related to execution, market-making and inventory management, offline conservative RL is suggested, adding CVaR or budget limits to constrain tail risks; Construct and evaluate the OPE confidence range and safety deployment criteria to assure controllability [41].

5.2.5. Search augmentation and safety boundaries for LLMs

Deploy RAG with re-ranking and answer with citations for LLMs answers and document generation. Thus, enhancing truth and trackability. Setting safety boundaries such as Jailbreaking and prompt implant protection, output audit to make the system from 'answerable' to 'truly trustworthy and auditable' [37,62].

5.2.6. Trends

Excluded from general LLMs, lite yet specialized domain models and agenting workflow cooperation are prioritized to reduce cost and latency while enhancing controllability. Regarding quantum computing, feasibility can be validated on subproblems such as portfolio optimization and sampling first, confirming its engineering bar and mid-term limitation while avoiding impractical expectations.

5.3. Governance and organization

Institutionalize model risk management: construct model catalog and tiering, unify document pattern, independent validation unit (IVU) and monitoring indicators, pushing challenger mechanisms and variation governance simultaneously. In terms of fairness, a fairness pipeline of

DOI: 10.54254/2755-2721/2025.LD29153

'data-model-output' can be constructed, along with the bias impact illustration and saving strategy (e.g., Heavy-weighted, threshold adjusting) [75,89]. Hence, forming a clear balance between individuals-group-business constraints. Clarify RACI obligation, setting human inspection, decision record and customer report channel for critical thresholds, generalizing human-machine cooperation.

5.4. Engineering and operation

Pushing Machine Learning operations (MLOps) and large language model operations (LLMOps) unification: Construct feature store and model registration [79,80], to enable canary release, automated roll-back by continuous integration (CI) and continuous delivery (CD) [82]. Deploying IAM and versioning to prompt and knowledge base. Regarding monitoring, apply multi-granularity linkage warning by business KPI, answer-rejection-rate, input feature distribution, calibration error etc. Forming a closed loop between orders and emergency plans [78,81]. In terms of performance and cost, utilize search-and-generate level caching, distillation and quantization, RAG freshness index; it is possible to gain a measurable balance between latency, cost and accuracy [37,56].

6. Conclusion

This article provides a structured review framework around AI in financial scenarios from technology evolution, method elements, business application and challenges. It emphasizes: on one hand, realized methods has demonstrated 'strong certainty' value among credit risk management, customer operation and regulatory technology; On the other hand, the real increase is the change from 'usable' to 'trustworthy'-interpretability, steadiness, compliance and engineering are the determinants of scale application. Looking ahead, paths illustrated at chapter five, including data governance, consequences learning, conservative RL etc., offer a practical blueprint for 'low risk, high auditability' productive AI. This article suggests adopting these elements in consideration of organizational lifespan and governance baseline, forming a reusable practice list. Hence, supporting the long-term and continuous intelligent upgrade to the financial industry.

Authorship statement

Hanlin Zhou and Junyi Huang are co-first authors of this paper.
Hanlin Zhou: Conceptualization, Investigation, Methodology
Junyi Huang: Writing, Review & editing

References

[1] Gould, M.D., Porter, M.A., Williams, S., McDonald, M., Fenn, D.J. and Howison, S.D., 2013. Limit order books. Quantitative Finance, 13(11), pp.1709-1742.
[2] Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M. and Bouchachia, A., 2014. A survey on concept drift adaptation. ACM computing surveys (CSUR), 46(4), pp.1-37.
[3] Bolton, R.J. and Hand, D.J., 2002. Statistical fraud detection: A review. Statistical science, 17(3), pp.235-255.
[4] Giarratano, J.C. and Riley, G., 1994. Expert systems: principles and programming. PWS Publishing Co..
[5] Breiman, L., 2001. Random forests. Machine learning, 45(1), pp.5-32.
[6] Chen, T. and Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
[7] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.
[8] Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J. and Williamson, R.C., 2001. Estimating the support of a high-dimensional distribution. Neural computation, 13(7), pp.1443-1471.

DOI: 10.54254/2755-2721/2025.LD29153

[9] Liu, F.T., Ting, K.M. and Zhou, Z.H., 2008, December. Isolation forest. In 2008 eighth ieee international conference on data mining (pp. 413-422). IEEE.
[10] Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S.A., Binder, A., Müller, E. and Kloft, M., 2018, July. Deep one-class classification. In International conference on machine learning (pp. 4393-4402). PMLR.
[11] Akoglu, L., Tong, H. and Koutra, D., 2015. Graph based anomaly detection and description: a survey. Data mining and knowledge discovery, 29(3), pp.626-688.
[12] Lee, D.H., 2013, June. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Workshop on challenges in representation learning, ICML (Vol. 3, No. 2, p. 896).
[13] Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A. and Li, C.L., 2020. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in neural information processing systems, 33, pp.596-608.
[14] Chen, T., Kornblith, S., Norouzi, M. and Hinton, G., 2020, November. A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607). PmLR.
[15] Arazo, E., Ortego, D., Albert, P., O'Connor, N.E. and McGuinness, K., 2020, July. Pseudo-labeling and confirmation bias in deep semi-supervised learning. In 2020 International joint conference on neural networks (IJCNN) (pp. 1-8). IEEE.
[16] Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R. and D'Oliveira, R.G., 2021. Advances and open problems in federated learning. Foundations and trends® in machine learning, 14(1–2), pp.1-210.
[17] Dwork, C. and Roth, A., 2014. The algorithmic foundations of differential privacy. Foundations and trends® in theoretical computer science, 9(3–4), pp.211-407.
[18] Bai, S., Kolter, J.Z. and Koltun, V., 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv: 1803.01271.
[19] Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. Neural computation, 9(8), pp.1735-1780.
[20] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv: 1406.1078.
[21] Lim, B., Arık, S.Ö., Loeff, N. and Pfister, T., 2021. Temporal fusion transformers for interpretable multi-horizon time series forecasting. International journal of forecasting, 37(4), pp.1748-1764.
[22] Gu, A., Johnson, I., Timalsina, A., Rudra, A. and Ré, C., 2022. How to train your hippo: State space models with generalized orthogonal basis projections. arXiv preprint arXiv: 2206.12037.
[23] Box, G.E., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M., 2015. Time series analysis: forecasting and control. John Wiley & Sons.
[24] Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31(3), pp.307-327.
[25] Zhang, G.P., 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, pp.159-175.
[26] Kipf, T.N., 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv: 1609.02907.
[27] Hamilton, W., Ying, Z. and Leskovec, J., 2017. Inductive representation learning on large graphs. Advances in neural information processing systems, 30.
[28] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P. and Bengio, Y., 2017. Graph attention networks. arXiv preprint arXiv: 1710.10903.
[29] Schlichtkrull, M., Kipf, T.N., Bloem, P., Van Den Berg, R., Titov, I. and Welling, M., 2018, June. Modeling relational data with graph convolutional networks. In European semantic web conference (pp. 593-607). Cham: Springer International Publishing.
[30] Veličković, P., Fedus, W., Hamilton, W.L., Liò, P., Bengio, Y. and Hjelm, R.D., 2018. Deep graph infomax. arXiv preprint arXiv: 1809.10341.
[31] Grover, A. and Leskovec, J., 2016, August. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 855-864).
[32] Perozzi, B., Al-Rfou, R. and Skiena, S., 2014, August. Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 701-710).
[33] Tao, Z. and Huang, J., 2024, December. Research on recommender systems based on GCN. In AIP Conference Proceedings (Vol. 3194, No. 1, p. 030017). AIP Publishing LLC.

DOI: 10.54254/2755-2721/2025.LD29153

[34] Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D. and Smith, N.A., 2020. Don't stop pretraining: Adapt language models to domains and tasks. arXiv preprint arXiv: 2004.10964.
[35] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A. and Schulman, J., 2022. Training language models to follow instructions with human feedback. Advances in neural information processing systems, 35, pp.27730-27744.
[36] Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L. and Chen, W., 2022. Lora: Low-rank adaptation of large language models. ICLR, 1(2), p.3.
[37] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.T., Rocktäschel, T. and Riedel, S., 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in neural information processing systems, 33, pp.9459-9474.
[38] Karpukhin, V., Oguz, B., Min, S., Lewis, P.S., Wu, L., Edunov, S., Chen, D. and Yih, W.T., 2020, November. Dense Passage Retrieval for Open-Domain Question Answering. In EMNLP (1) (pp. 6769-6781).
[39] Nogueira, R. and Cho, K., 2019. Passage Re-ranking with BERT. arXiv preprint arXiv: 1901.04085.
[40] Sutton, R.S., McAllester, D., Singh, S. and Mansour, Y., 1999. Policy gradient methods for reinforcement learning with function approximation. Advances in neural information processing systems, 12.
[41] Levine, S., Kumar, A., Tucker, G. and Fu, J., 2020. Offline reinforcement learning: Tutorial, review, and perspectives on open problems. arXiv preprint arXiv: 2005.01643.
[42] Fujimoto, S., Meger, D. and Precup, D., 2019, May. Off-policy deep reinforcement learning without exploration. In International conference on machine learning (pp. 2052-2062). PMLR.
[43] Chow, Y. and Ghavamzadeh, M., 2014. Algorithms for CVaR optimization in MDPs. Advances in neural information processing systems, 27.
[44] Tetlock, P.C., 2007. Giving content to investor sentiment: The role of media in the stock market. The Journal of finance, 62(3), pp.1139-1168.
[45] Markowitz, H. 1952, 'Portfolio selection', The Journal of Finance, vol. 7, no. 1, pp. 77–91, doi: 10.2307/2975974.
[46] Almgren, R. and Chriss, N., 2001. Optimal execution of portfolio transactions. Journal of Risk, 3, pp.5-40.
[47] Avellaneda, M. and Stoikov, S., 2008. High-frequency trading in a limit order book. Quantitative Finance, 8(3), pp.217-224.
[48] Bailey, D.H. and López de Prado, M., 2014. The deflated Sharpe ratio: Correcting for selection bias, backtest overfitting and non-normality. Journal of Portfolio Management, 40(5), pp.94-107.
[49] De Prado, M.L., 2018. Advances in financial machine learning. John Wiley & Sons.
[50] Bailey, D.H. and Lopez de Prado, M., 2012. The Sharpe ratio efficient frontier. Journal of Risk, 15(2), p.13.
[51] Bailey, D.H., Borwein, J.M., De Prado, M.L. and Zhu, Q.J., 2014. Pseudomathematics and financial charlatanism: The effects of backtest over fitting on out-of-sample performance. Notices of the AMS, 61(5), pp.458-471.
[52] White, H., 2000. A reality check for data snooping. Econometrica, 68(5), pp.1097-1126.
[53] Thomas, L., Crook, J. and Edelman, D., 2017. Credit scoring and its applications. Society for industrial and Applied Mathematics.
[54] Radcliffe, N.J. and Surry, P.D., 2011. Real-world uplift modelling with significance-based uplift trees. White Paper TR-2011-1, Stochastic Solutions, pp.1-33.
[55] Basel Committee on Banking Supervision (2019) Minimum capital requirements for market risk (January 2019; corrected February 2019). Bank for International Settlements. Available at: https: //www.bis.org/bcbs/publ/d457.htm (Accessed: 14 September 2025).
[56] Saito, T. and Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PloS one, 10(3), p.e0118432.
[57] Hinton, G., Vinyals, O. and Dean, J., 2015. Distilling the knowledge in a neural network. arXiv preprint arXiv: 1503.02531.
[58] Caruana, R., 1997. Multitask learning. Machine learning, 28(1), pp.41-75.
[59] Auer, P., Cesa-Bianchi, N. and Fischer, P., 2002. Finite-time analysis of the multiarmed bandit problem. Machine learning, 47(2), pp.235-256.
[60] Lundberg, S.M. and Lee, S.I., 2017. A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.
[61] Wachter, S., Mittelstadt, B. and Russell, C., 2017. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. Harv. JL & Tech., 31, p.841.
[62] Financial Action Task Force (FATF) 2020, Guidance on Digital Identity, FATF/OECD, Paris. Available at: https: //www.fatf-gafi.org/publications/documents/digital-identity-guidance.html (Accessed 14 September 2025).
[63] Arner, D.W., Barberis, J. and Buckey, R.P., 2016. FinTech, RegTech, and the reconceptualization of financial regulation. Nw. J. Int'l L. & Bus., 37, p.371.

DOI: 10.54254/2755-2721/2025.LD29153

[64] National Institute of Standards and Technology (NIST). 2023. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. Gaithersburg, MD: U.S. Department of Commerce. Available at: https: //doi.org/10.6028/NIST.AI.100-1 (Accessed 14 September 2025).
[65] Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.F. and Dennison, D., 2015. Hidden technical debt in machine learning systems. Advances in neural information processing systems, 28.
[66] Settles, B., 2009. Active learning literature survey.
[67] Ricci, F., Rokach, L. and Shapira, B., 2010. Introduction to recommender systems handbook. In Recommender systems handbook (pp. 1-35). Boston, MA: springer US.
[68] Wang, R.Y. and Strong, D.M., 1996. Beyond accuracy: What data quality means to data consumers. Journal of management information systems, 12(4), pp.5-33.
[69] Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature machine intelligence, 1(5), pp.206-215.
[70] Hendrycks, D. and Gimpel, K., 2016. A baseline for detecting misclassified and out-of-distribution examples in neural networks. arXiv preprint arXiv: 1610.02136.
[71] Biggio, B., Nelson, B. and Laskov, P., 2012. Poisoning attacks against support vector machines. arXiv preprint arXiv: 1206.6389.
[72] Guo, C., Pleiss, G., Sun, Y. and Weinberger, K.Q., 2017, July. On calibration of modern neural networks. In International conference on machine learning (pp. 1321-1330). PMLR.
[73] Kaufman, S., Rosset, S., Perlich, C. and Stitelman, O., 2012. Leakage in data mining: Formulation, detection, and avoidance. ACM Transactions on Knowledge Discovery from Data (TKDD), 6(4), pp.1-21.
[74] Thomas, P. and Brunskill, E., 2016, June. Data-efficient off-policy policy evaluation for reinforcement learning. In International conference on machine learning (pp. 2139-2148). PMLR.
[75] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y.J., Madotto, A. and Fung, P., 2023. Survey of hallucination in natural language generation. ACM computing surveys, 55(12), pp.1-38.
[76] Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T. and Fritz, M., 2023, November. Not what you've signed up for: Compromising real-world llm-integrated applications with indirect prompt injection. In Proceedings of the 16th ACM workshop on artificial intelligence and security (pp. 79-90).
[77] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A., 2021. A survey on bias and fairness in machine learning. ACM computing surveys (CSUR), 54(6), pp.1-35.
[78] European Union 2016, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation), consolidated text available at GDPR-Info.eu. Available at: https: //gdpr-info.eu/ (Accessed 14 September 2025).
[79] Crankshaw, D., Wang, X., Zhou, G., Franklin, M.J., Gonzalez, J.E. and Stoica, I., 2017. Clipper: A {Low-Latency} online prediction serving system. In 14th USENIX Symposium on Networked Systems Design and Implementation (NSDI 17) (pp. 613-627).
[80] Breck, E., Cai, S., Nielsen, E., Salib, M. and Sculley, D., 2017, December. The ML test score: A rubric for ML production readiness and technical debt reduction. In 2017 IEEE international conference on big data (big data) (pp. 1123-1132). IEEE.
[81] Baylor, D., Breck, E., Cheng, H.T., Fiedel, N., Foo, C.Y., Haque, Z., Haykal, S., Ispir, M., Jain, V., Koc, L. and Koo, C.Y., 2017, August. Tfx: A tensorflow-based production-scale machine learning platform. In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1387-1395).
[82] Zaharia, M., Chen, A., Davidson, A., Ghodsi, A., Hong, S.A., Konwinski, A., Murching, S., Nykodym, T., Ogilvie, P., Parkhe, M. and Xie, F., 2018. Accelerating the machine learning lifecycle with MLflow. IEEE Data Eng. Bull., 41(4), pp.39-45.
[83] Polyzotis, N., Roy, S., Whang, S.E. and Zinkevich, M., 2017, May. Data management challenges in production machine learning. In Proceedings of the 2017 ACM international conference on management of data (pp. 1723-1726).
[84] Beyer, B., Jones, C., Petoff, J. and Murphy, N.R., 2016. Site reliability engineering: how Google runs production systems. " O'Reilly Media, Inc.".
[85] Buneman, P., Khanna, S. and Wang-Chiew, T., 2001, January. Why and where: A characterization of data provenance. In International conference on database theory (pp. 316-330). Berlin, Heidelberg: Springer Berlin Heidelberg.
[86] Rosenbaum, P.R. and Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), pp.41-55.

DOI: 10.54254/2755-2721/2025.LD29153

[87] Sagawa, S., Koh, P. W., Hashimoto, T. B., & Liang, P. (2019). Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. arXiv preprint arXiv: 1911.08731.
[88] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. and Robins, J., 2018. Double/debiased machine learning for treatment and structural parameters.
[89] Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D. and Gebru, T., 2019, January. Model cards for model reporting. In Proceedings of the conference on fairness, accountability, and transparency (pp. 220-229).