

# A transferability benchmark of graph neural networks for cross-domain risk contagion identification: GAT, GraphSAGE, and GCN across financial, cyber, and epidemiological networks

Daniel R. Whitaker

Department of Computational Risk Analytics, College of Engineering, Northeastern University, Boston, MA 02115, USA

## Abstract

Risk contagion across heterogeneous networks—shock propagation between financial markets, lateral movement of intrusions across enterprise hosts, and pathogen spreading along contact graphs—shares a common formal structure but is rarely studied jointly. We construct a unified benchmark that aligns three publicly derived graph corpora (an interbank-style transaction subgraph, an authentication-flow subgraph from enterprise telemetry, and an SNAP contact graph augmented with synthetic exposure labels) under a single node-classification protocol, and run a controlled comparison of three widely deployed graph neural networks: Graph Convolutional Networks (GCN), GraphSAGE, and Graph Attention Networks (GAT). Models are evaluated in three regimes: in-domain training, cold-start cross-domain transfer with frozen encoders, and adapter-based fine-tuning on a 5% target sample. Across 27 (model  $\times$  source  $\times$  target) configurations, GAT attains the highest in-domain F1 and the most stable transfer in financial $\rightarrow$ cyber pairs, GraphSAGE leads on the epidemiological target due to neighborhood-sampling robustness on long-tailed degree distributions, and GCN shows the steepest performance gap when transferred without adaptation. We discuss when neighborhood attention helps versus when sampling-based aggregation suffices, and present an analysis of feature drift and label-noise sensitivity. The benchmark, while not exhaustive, offers a reproducible reference point for cross-domain GNN deployment in risk-sensitive applications.

**Keywords:** Graph neural networks; risk contagion; cross-market financial spillover; lateral movement detection; epidemic spreading; transfer learning; benchmark.

## 1. Introduction

Risk seldom respects domain boundaries. A liquidity stress event in one corner of an equity market can cascade across asset classes within minutes<sup>[1]</sup>; a single compromised credential triggers lateral pivots that expose adjacent hosts before any signature-based defense reacts<sup>[2]</sup>; a respiratory pathogen introduced into one community ward radiates through contact patterns that look, structurally, very similar to the previous two<sup>[3]</sup>. Recent empirical work has emphasized that the abstract object underlying all three is a graph in which a small set of "amplifier" nodes carries an outsized share of contagion mass<sup>[4]</sup>, and that identifying those amplifiers ahead of time is more valuable than reconstructing the cascade after the fact<sup>[5]</sup>.

Despite this conceptual convergence, the literature on risk contagion has remained largely siloed. Financial-economics research models cross-market spillover with VAR-derived networks and centrality measures, sometimes augmented with credit-transmission analyses for supply-chain finance<sup>[6]</sup> or with multi-source data fusion to predict consumer-credit defaults<sup>[7]</sup>. Cybersecurity work approaches lateral movement through attack-path reasoning<sup>[8]</sup>, graph-based behavioral detection of supply-chain intrusions<sup>[9]</sup>, and ensemble learning over temporal threat features<sup>[2]</sup>. Epidemiological modeling has, since 2020, increasingly drawn on machine-learning hospital-utilization forecasts during outbreak surges<sup>[10]</sup> and rule-enhanced risk stratification for vulnerable elderly cohorts<sup>[11]</sup>. The methods used in each silo overlap substantially—node embeddings, message passing, attention—yet head-to-head transferability evidence is scarce.

Graph neural networks (GNNs) are an obvious bridging tool. GCN, GraphSAGE, and GAT have all been deployed for financial fraud and anomalous transactions<sup>[12]</sup>, for malicious-account detection on social platforms via temporal graph features<sup>[13, 14]</sup>, and—in adjacent biomedical contexts—for graph-attention-based feature selection on multi-omics data<sup>[15]</sup>. Knowledge-graph completion methods have likewise pushed relational message passing toward higher representational efficiency<sup>[16]</sup>. What is missing is a controlled comparison: when the same three architectures are presented with structurally analogous but semantically different graphs, which inductive bias actually transfers, and where do the gains come from?

This paper provides such a comparison. Our contributions are:

(1) **Three aligned subtasks under one protocol.** We assemble a financial cross-market subgraph (anonymized inter-institution exposure flows), a cyber lateral-movement subgraph (authentication and

process-launch edges from enterprise host telemetry, in the spirit of [17]), and an epidemiological contact subgraph (SNAP-derived contact networks augmented with synthetic exposure labels), and convert each into a binary node-classification task with a calibrated positive rate.

(2) **A controlled comparison of GCN, GraphSAGE, and GAT** under three regimes: in-domain training, cold-start cross-domain transfer with frozen encoders, and adapter-based fine-tuning on a 5% target sample. The third regime echoes the spirit of cross-context transfer-learning evaluation in behavioral generalization [18].

(3) **An empirical map of where each architecture wins.** GAT dominates in-domain financial scoring; GraphSAGE leads under degree-skewed contact graphs; GCN remains a competitive baseline only when the source and target graphs share comparable density.

(4) **A discussion of operational constraints**—privacy, fairness, and robustness—that determine whether a model that wins in benchmarks is actually deployable in regulated environments such as community banks [19], cross-border financial transaction monitoring [20], or healthcare claims surveillance [21].

The remainder of the paper is organized as follows. Section 2 reviews the three application literatures and the GNN methods that have grown out of them. Section 3 formalizes the cross-domain task and details the three architectures. Section 4 describes the dataset construction and evaluation protocol. Section 5 reports results. Section 6 discusses operational and ethical implications. Section 7 concludes.

## 2. Related Work

### 2.1. Risk contagion in financial networks

Empirical analyses of US credit-equity stress periods document risk-amplifier nodes whose pre-shock topology already encodes their post-shock importance [5]. Cross-asset liquidity contagion pathways have been recovered with AI-enhanced hedging-strategy optimization [1], and graph-attention networks have been applied directly to identify cross-market amplifiers [4]. At the institutional level, lightweight stress testing using variational autoencoders combined with extreme value theory has been proposed for small and medium financial institutions [22], and ensemble anomaly detection with explainable AI has been deployed for community-bank multi-risk warning [19]. Adjacent work on counterparty exposure includes deep-learning dynamic margin-period-of-risk prediction [23] and adaptive importance sampling for jump-diffusion CVA [24]. Synthetic-identity fraud and money-laundering detection in real time has likewise been tackled through enhanced feature engineering [25], with related contributions on AML compliance auditing [26], supply-chain credit transmission [6], and intelligent credit-risk assessment for small and medium enterprises based on multi-dimensional data fusion [27].

Beyond bilateral exposure, market-text channels matter. NLP-driven risk-signal detection on asset-backed securities text [28] and ESG news sentiment quantification for portfolio outcomes [29] have shown that off-balance-sheet narrative content carries non-trivial predictive content for systemic risk. Risk-aware portfolio optimization for ultra-high-net-worth clients in volatile markets [30] and dynamic asset allocation for target-date funds [31] connect these signals back to portfolio actions. Behavioral-sequence-based real-time fraud scoring [32] and time-decay-aware incremental feature extraction for transaction fraud [33] further illustrate that contagion-relevant features are themselves time-varying.

### 2.2. Lateral movement and cyber contagion

Lateral movement has been modeled as a graph-traversal problem under uncertain edge weights. Industrial-control attack-path reasoning over knowledge graphs offers explainable hop-by-hop scoring [8], while supply-chain intrusion detection has been recast as a graph-learning problem over software-package dependency edges [9]. AI-driven temporal threat classification [2] complements ML-based risk frameworks for data-leakage prevention [34] and federated-learning-based privacy-preserving network analytics [35, 36]. Firmware vulnerability detection with hybrid analysis [37] and zero-day attack early warning in cloud infrastructure [38] cover the host- and infrastructure-side of the same problem, and telemetry-driven detection has been compared specifically for dual-purpose operational and security optimization at the edge [39]. Anomalous login behavior in enterprise networks [17] and large-language-model-assisted threat intelligence [40] round out the toolkit.

### 2.3. Epidemic-style contagion on contact graphs

Public-health early-warning systems combine spatial mobility, contact reports, and environmental signals into community-scale infection forecasts [3]. Hospital resource demand under infectious-disease surges has been benchmarked against ML forecasters [10]. Risk stratification for adverse outcomes in elderly cohorts—e.g., polypharmacy interactions [11]—and integrated prediction-intervention-evaluation frameworks for hospital readmission management [41] use graph-like patient-similarity structures even when not labeled as graphs. Healthcare claims temporal anomaly detection [21] and intelligent recognition of anomalous medical-insurance

behavior<sup>[42]</sup> use techniques structurally close to those used for financial transaction anomalies—an early hint that cross-domain transfer might be feasible.

## 2.4. GNN architectures and methodological neighbors

The three GNN families studied here—GCN (spectral neighborhood averaging), GraphSAGE (sampled-neighborhood aggregation), and GAT (attention-weighted aggregation)—have each accumulated domain-specific evidence. Graph-based representation learning has become a default for financial-fraud and anomaly-transaction detection<sup>[12]</sup>. Graph-attention-based feature selection in multi-omics drug-target prediction<sup>[15]</sup> illustrates the same attention bias that we will probe here on financial graphs. Temporal graph feature learning for early detection of malicious accounts<sup>[13]</sup> and graph-based temporal behavior analysis for coordinated malicious accounts<sup>[14]</sup> confirm that the time axis matters for both cyber and social-platform variants of lateral movement. Knowledge-graph completion through efficient relational context perception<sup>[16]</sup> sits methodologically adjacent: it shares the message-passing formalism but places more weight on edge typing than node features.

Comparative studies of ML algorithms on structured tasks—e.g., filter-based feature-selection methods in high-dimensional classification<sup>[43]</sup> and oversampling-ensemble interactions on imbalanced tabular data<sup>[44]</sup>—provide the methodological hygiene needed for fair benchmarking. Multi-source data fusion for credit-default early warning<sup>[7]</sup> and feature-attribution explainability for market-risk stress scenarios<sup>[45]</sup> feed directly into the feature-engineering and explainability components of our pipeline.

## 2.5. Cross-domain transfer and operational constraints

Whether a model trained on one risk graph generalizes to another is partly a transfer-learning question. Comparative evaluation of transfer learning for cross-context behavioral generalization in autism-spectrum interventions<sup>[18]</sup> is technically far from finance, but the protocol it uses—source–target performance gap, adapter fine-tuning, holdout context—maps cleanly onto our setting. Agentic AI surveys note that practical deployments increasingly compose specialized models across heterogeneous tasks rather than retraining single monoliths<sup>[46]</sup>, a trend echoed in multi-agent online learning under dynamic environments<sup>[47]</sup> and web-agent reinforcement learning under multi-cost and failure-risk constraints<sup>[48]</sup>. These pieces motivate the engineering choice we make in Section 3 to factor encoders from heads.

Operational constraints—privacy, fairness, robustness—are first-order in deployed risk systems. Privacy-preserving collaborative learning across healthcare institutions<sup>[49]</sup>, adaptive privacy-budget allocation for federated learning in healthcare<sup>[50]</sup>, and a systematic review of privacy-preserving federated learning in medical AI<sup>[51]</sup> map directly onto the cross-institutional settings we anticipate for the financial and epidemiological subtasks. Differential-privacy mechanisms for federated document classification<sup>[52]</sup>, collaborative privacy-preserving risk monitoring across financial institutions<sup>[53]</sup>, and privacy-aware AI for rare-disease patient discovery<sup>[54]</sup> complete this picture. Fairness-aware feature attribution for credit scoring<sup>[55]</sup> and fairness-accuracy trade-offs in AI credit scoring under multiple constraints<sup>[56]</sup>, together with fairness-aware multimodal fusion for chronic-disease risk prediction<sup>[57]</sup>, will reappear in our discussion as constraints on model selection.

A summary of the most directly comparable studies appears in Table 1.

Table 1. Summary of related studies and their technical contributions referenced in this benchmark.

Study (cite)	Domain	Graph / Method	Key contribution to our framework
<sup>[3]</sup>	Finance	Network topology analysis	Pre-shock topological signatures of stress
<sup>[1]</sup>	Finance	AI-enhanced contagion mapping	Cross-asset liquidity pathways and hedging
<sup>[4]</sup>	Finance	Graph attention network	Cross-market amplifier identification
<sup>[12]</sup>	Finance	Graph representation learning	Fraud / anomaly transaction baseline

Study (cite)	Domain	Graph / Method	Key contribution to our framework
[6]	Finance	Credit-transmission analysis	Supply-chain credit propagation prior
[2]	Cyber	Ensemble + temporal features	Threat-pattern classification baseline
[8]	Cyber	Knowledge-graph reasoning	Explainable attack-path scoring
[9]	Cyber	Graph learning	Software supply-chain intrusion detection
[17]	Cyber	ML comparison	Anomalous-login labeling protocol
[3]	Public health	Practical AI early warning	Community-infection labeling reference
[10]	Public health	ML forecasting	Hospital-resource demand under surges
[15]	Biomedical	Graph attention feature selection	Attention bias evidence transferable
[16]	KG completion	Relational message passing	Edge-type-aware aggregation reference
[18]	Behavioral	Transfer-learning evaluation	Cross-context fine-tuning protocol
[7]	Finance	Multi-source fusion	Credit-default early warning features

### 3. Methodology

#### 3.1. Problem formulation

Let  $G = (V, E, X)$  be a directed attributed graph, with node features  $X \in \mathbb{R}^{(|V| \times d)}$  and a binary label vector  $y \in \{0, 1\}^{|V|}$  where  $y_i = 1$  marks node  $i$  as a *contagion-positive* node. The semantics of "contagion-positive" differ by subtask: a financial node is positive if it experiences a meaningful exposure shock within a fixed window after a market stress event; a host is positive if it appears as an intermediate or terminal node on a confirmed lateral-movement path; and a person/contact-node is positive if it tests symptomatic or seroconverts within an incubation window. The *cross-domain* task asks: train a model on a source graph  $G^s$  and predict  $y^t$  on a target graph  $G^t$  from a different domain.

We focus on three regimes. **Regime A (in-domain)**: train and evaluate on the same domain, with a temporal split. **Regime B (zero-shot transfer)**: freeze the encoder trained on  $G^s$  and attach a randomly initialized linear head on  $G^t$ , training only the head. **Regime C (5%-fine-tune)**: as B, but allow the last GNN layer plus the head to adapt on a 5% target sample.

#### 3.2. Three GNN architectures

We study three architectures with the message-passing form  $h_v^{(l+1)} = \sigma(W^{(l)} \cdot \text{AGG}(\{h_u^{(l)} : u \in N(v)\}) + b^{(l)})$ .

**GCN.** Aggregation is symmetric normalized:  $AGG(\cdot) = \tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} H$ . We use 2 layers with hidden dimension 128 and ReLU.

**GraphSAGE.** Neighborhoods are sampled (sizes 25 and 10 at the two layers), and aggregation is mean-pooled with concatenation of self and neighborhood features. This sampling is what we expect to help on heavy-tailed degree distributions characteristic of social-contact and authentication graphs, an inductive bias whose value has also been observed for clustering-based banking customer segmentation [58].

**GAT.** Attention weights  $\alpha_{vu}$  are learned per edge with a single-head LeakyReLU score, and we use 4 heads at the first layer (concatenated) and 1 head at the second (averaged). Following the attention-bias evidence from biomedical multi-omics [15], we expect attention to help most where node-feature signal-to-noise is variable—plausibly on the financial subtask, where some institutional nodes are sparsely featurized.

A side-by-side comparison of the three families on dimensions that matter for risk applications appears in Table 2.

Table 2. Comparison of the three GNN architectures across properties relevant to risk-contagion graphs.

Property	GCN	GraphSAGE	GAT
Aggregation	Symmetric normalized mean	Sampled-neighborhood mean (with concat)	Attention-weighted sum
Per-edge weighting	Fixed adjacency	by Uniform within sample	Learned per edge
Scalability to dense hubs	Moderate	High (sampling)	Lower (full attention)
Robustness to label noise	Moderate	Higher empirically	Lower at high flip rate
Best fit (this study)	Sparse, homogeneous graphs	Heavy-tailed degree, large hubs	Variable per-node SNR
Hyperparameters tuned	Hidden dim; layers	Sample sizes per layer	Heads per layer; head averaging
Computational cost (1 epoch, FIN)	0.31 s	0.42 s	0.78 s

### 3.3. Feature engineering

Each subtask carries a base feature set augmented by domain-specific signals. For the financial graph we include rolling-window return volatility, exposure share, and a risk-text-embedding feature derived from quarterly disclosure language—the latter a concession to text-mining evidence in asset-backed-securities risk monitoring [28] and to recent NLP work on contingent-liability classification [59]. For the cyber graph we include host metadata, behavioral counters (logon hour entropy, session duration), and a suspicion score from a separate LSTM-based behavioral pipeline broadly aligned with [32]. For the contact graph we include demographic group indicators, average daily contact count, and a household-membership indicator derived from co-location runs.

Feature pre-selection follows the comparative practice surveyed in [43] for high-dimensional classification, and we apply class-imbalance correction following the oversampling-ensemble guidance in [44]. Where required for institutional-data privacy we use differentially private gradient updates following the budget-allocation strategy in [50].

### 3.4. Cross-domain transfer protocol

The encoder  $f_\theta$  produces a 64-dimensional node embedding. In Regime B, we attach a head  $g_\phi$  trained only on a small labeled subset of  $G^*(t)$ . In Regime C, the last GNN layer's parameters  $\theta_L$  are unfrozen and trained jointly with  $\phi$ . Domain-shift sensitivity is monitored with two diagnostics: a feature-drift score (Maximum Mean Discrepancy on encoder activations) and a label-noise probe in which a fraction of  $y^*(s)$  is flipped before transfer, an experiment design motivated by recent findings on memory-poisoning propagation in multi-agent systems [60]. Where domain-specific labels are sparse, semi-supervised pretraining on the target graph uses the contrastive-learning pattern that has been effective for label-efficient medical imaging [61].

The overall research framework is depicted in Figure 1.

[Figure 1 — Overall research framework. Three parallel data ingestion lanes (FIN: inter-institution exposure flows; CYB: authentication telemetry; EPI: SNAP contact graph) feed into a shared graph-construction module that produces a directed/undirected attributed graph  $G = (V, E, X)$ . Encoders  $f_\theta \in \{GCN, GraphSAGE, GAT\}$  produce 64-dim node embeddings. A switch routes embeddings to one of three regimes: in-domain training (Regime A), zero-shot transfer with frozen encoder (Regime B), or 5%-fine-tune adaptation (Regime C). Output passes through evaluation block reporting F1, AUPRC, ECE, and contagion lift, plus fairness audit and drift diagnostic (MMD, label-noise probe).]

Figure 1. Overall research framework: data ingestion across three risk domains, shared graph construction, three GNN encoders, three transfer regimes, evaluation and audit blocks.

A more detailed view of the per-step methodological pipeline—feature ingestion, graph construction, encoder training, transfer protocol, and evaluation—is shown in Figure 2.

[Figure 2 — Methodological pipeline. Step 1: domain-specific feature engineering (volatility/exposure/text for FIN; behavior counters/suspicion score for CYB; demographics/contact percentile for EPI). Step 2: graph construction with temporal split. Step 3: encoder pre-training under Regime A. Step 4: cross-domain transfer (Regime B = freeze + linear head; Regime C = unfreeze last layer + head). Step 5: evaluation pipeline producing per-seed F1/AUPRC/ECE plus MMD-based feature-drift diagnostics. Arrows annotated with the privacy guard (DP/federated) inserted between Steps 2 and 3 in cross-institutional deployment.]

Figure 2. Detailed methodological pipeline showing the per-step flow from domain-specific feature engineering through encoder training and cross-domain transfer to evaluation.

## 4. Experimental Design and Data Construction

### 4.1. Data sources

The three subgraphs are constructed as follows.

**Financial subgraph (FIN).** We use a synthetic but calibration-realistic inter-institution exposure graph with 2,418 nodes and 39,704 directed edges. Edge weights encode counterparty exposure normalized by the smaller of the two institutions' regulatory capital. The construction logic follows guidance for AI-enhanced cross-asset liquidity analysis [1] and uses a market-stress event window as the labeling trigger; positive class prevalence is 8.7%. Risk-text embeddings are produced with a sentence-transformer encoder. Where supplemental signals are required—e.g., XBRL-tag semantic-mismatch flags drawn from SEC 10-K analysis [62] or disclosure-discrepancy markers [63]—they are folded into the node feature vector. Cross-border anomalous fund-flow features, when available, are added following [20].

**Cyber subgraph (CYB).** We use an authentication-flow subgraph drawn from a public enterprise telemetry replay (12,776 host nodes, 81,553 directed edges). Lateral-movement labels are derived from the published incident annotations with the temporal-graph-feature methodology in [13] and the malicious-account-coordination analysis in [14]; positive prevalence is 4.1%. Host-level behavioral features include process-launch entropy and credential-reuse counters, with a derived "explainable-attack-path" priority score in the spirit of [8]. We also incorporate a zero-day detection prior derived from anomaly-based detection literature [38] and a telemetry-effectiveness baseline aligned with [39].

**Epidemiological subgraph (EPI).** We use a SNAP contact graph (a co-location network derived from a public mobility dataset, 7,624 nodes and 27,806 undirected edges). Exposure labels are simulated under an SEIR-style process calibrated against community-infection-warning evidence [3]; positive prevalence is 6.3%. Node features include demographics, contact-count percentiles, and a household-membership flag.

For each graph we construct a temporal split (60% / 20% / 20%) so that test-time edges and labels are strictly later than training-time ones. The split is necessary because cross-sectional shuffles can produce optimistic estimates of contagion-prediction skill [64, 65].

## 4.2. Scenario configurations

Beyond the three within-domain trainings, we run six cross-domain transfer pairs and three "domain-mixture" trainings (where two source domains are concatenated). Table 3 summarizes the experimental configurations and the size of each split.

Table 3. Experimental dataset and scenario configuration. For transfer rows, "src" indicates labeled source-domain training nodes and "tgt" indicates the 5% target adaptation sample used in Regime C.

Subtask	V	E	Pos. rate	Train / val / test (nodes)	Edge type
FIN (financial)	2,418	39,704	8.7%	1,451 / 484 / 483	Directed weighted
CYB (cyber)	12,776	81,553	4.1%	7,666 / 2,555 / 2,555	Directed
EPI (contact)	7,624	27,806	6.3%	4,575 / 1,524 / 1,525	Undirected
FIN→CYB transfer	— / 12,776	— / 81,553	— / 4.1%	src 1,451 + tgt 638 / val 1,277 / test 2,555	Mixed
CYB→FIN transfer	— / 2,418	— / 39,704	— / 8.7%	src 7,666 + tgt 121 / val 242 / test 483	Mixed
FIN→EPI transfer	— / 7,624	— / 27,806	— / 6.3%	src 1,451 + tgt 381 / val 762 / test 1,525	Mixed
EPI→FIN transfer	— / 2,418	— / 39,704	— / 8.7%	src 4,575 + tgt 121 / val 242 / test 483	Mixed
CYB→EPI transfer	— / 7,624	— / 27,806	— / 6.3%	src 7,666 + tgt 381 / val 762 / test 1,525	Mixed
EPI→CYB transfer	— / 12,776	— / 81,553	— / 4.1%	src 4,575 + tgt 638 / val 1,277 / test 2,555	Mixed

## 4.3. Baselines

We compare against four non-GNN baselines: logistic regression on a hand-engineered tabular feature set; XGBoost with structural-statistic features (degree, k-core, PageRank); a 2-layer multilayer perceptron on the same tabular set; and a node2vec embedding plus logistic head. The XGBoost baseline is informed by ensemble-comparison evidence in online payment-fraud detection<sup>[66]</sup> and by ensemble-anomaly-detection deployment in community-bank multi-risk warning<sup>[19]</sup>. The node2vec baseline is included to isolate the contribution of message passing beyond random-walk-based embeddings.

## 4.4. Metrics

We report: F1 at the operating point with maximum F1 on validation; Precision@k where k is calibrated to the observed positive prevalence on the test split; AUPRC (area under the precision–recall curve), preferred over AUROC for imbalanced settings; and a *contagion lift* metric that compares predicted top-decile contagion probability to the realized post-event spread, computed analogously to the lead-time evaluation in<sup>[67]</sup>. Confidence calibration is reported using the Expected Calibration Error (ECE), motivated by the trustworthiness-evaluation framework in<sup>[68]</sup>. For fairness checks (Section 6) we compute prevalence parity across protected subgroups, drawing on the methodology of fairness-accuracy trade-off evaluation<sup>[56]</sup> and fairness-aware feature attribution<sup>[55]</sup>.

Implementation uses PyTorch Geometric. Training runs on a single NVIDIA A100 (40 GB) with mixed precision; each model was trained 5 times under different seeds, and reported numbers are means with 95% bootstrap confidence intervals.

## 5. Results and Analysis

### 5.1. In-domain results

Table 4 reports test-set F1, AUPRC, and ECE for each (architecture, domain) pair, as well as the four baselines.

Table 4. Performance comparison across architectures and transfer regimes. Cell entries are mean F1 over 5 seeds. Two-value entries report (Regime B / Regime C).

Model	FIN F1	CYB F1	EPI F1	FIN→CYB (B/C)	CYB→FIN (B/C)	FIN→EPI (B/C)	EPI→FIN (B/C)	AUPRC (FIN)
LogReg	0.518	0.487	0.451	— / —	— / —	— / —	— / —	0.421
XGBoost	0.668	0.652	0.581	— / —	— / —	— / —	— / —	0.594
MLP	0.602	0.561	0.512	— / —	— / —	— / —	— / —	0.508
node2vec+LR	0.629	0.596	0.527	0.241 0.412	/ 0.236 0.408	/ 0.198 0.357	/ 0.184 0.351	/ 0.547
GCN	0.692	0.640	0.604	0.276 0.523	/ 0.281 0.541	/ 0.221 0.413	/ 0.207 0.397	/ 0.611
GraphSAGE	0.728	0.683	0.622	0.312 0.581	/ 0.327 0.598	/ 0.265 0.487	/ 0.249 0.451	/ 0.652
GAT	0.741	0.671	0.609	0.301 0.594	/ 0.318 0.572	/ 0.234 0.461	/ 0.218 0.428	/ 0.673

In-domain, GAT obtains the best F1 on the financial subtask (0.741, 95% CI [0.728, 0.755]), narrowly beating GraphSAGE (0.728) and GCN (0.692). The gap is mostly driven by recall on mid-degree institutional nodes whose feature vectors are sparse: attention appears to compensate by upweighting a small set of informative neighbors, consistent with the multi-omics observation in <sup>[15]</sup>. On the cyber subtask, GraphSAGE leads (F1 0.683, vs. GAT 0.671 and GCN 0.640); we attribute this to the heavy-tailed authentication-graph degree distribution, where GraphSAGE's neighborhood sampling reduces noise from hub nodes that connect to thousands of low-information peers. On the epidemiological subtask, GraphSAGE leads again (F1 0.622), likely for the same degree-distribution reason. GCN remains a tight third on EPI (0.604), reflecting that symmetric averaging is not catastrophic when the graph is comparatively sparse and homogeneous.

The XGBoost baseline outperforms GCN on the cyber subtask, which is consistent with prior findings that strong ensemble methods on hand-engineered structural features can rival message-passing networks under certain conditions <sup>[66]</sup>. Node2vec is uniformly weaker than the GNNs but is competitive with the MLP on FIN and EPI.

### 5.2. Cross-domain transfer

Table 4 also reports F1 under Regime B (zero-shot) and Regime C (5% fine-tune) for the six (source, target) pairs. Two patterns are visible.

First, the FIN→CYB and CYB→FIN transfer pairs are the most successful. Under Regime C, GAT trained on FIN reaches an F1 of 0.594 on CYB; GraphSAGE trained on CYB reaches 0.598 on FIN. Both are within 12–14 percentage points of in-domain performance—a non-trivial gap, but small enough that the encoders can plausibly be reused with adapter fine-tuning. The structural reason, we suspect, is that interbank-exposure flows and authentication flows share a "few-hub-many-leaves" shape that the GNNs implicitly exploit. This echoes findings on cross-asset contagion-pathway identification <sup>[1]</sup> that the topological signature of a stress event is partially domain-invariant.

Second, transfer involving EPI is uniformly harder. Under Regime B, all three architectures collapse below 0.30 F1 in either direction. Under Regime C, the gap narrows—GraphSAGE recovers to 0.487 on FIN→EPI—but the improvement requires more than the 5% target sample we permit, and the variance across seeds is

substantial. The contact graph's undirected, sparse, and demographically heterogeneous character is genuinely different from the directed, dense, hub-dominated financial and cyber graphs.

A bar-chart visualization of relative performance across the nine in-domain  $\times$  cross-domain configurations is provided in Figure 3.

[Figure 3 — Performance comparison bar chart. Grouped bars (one group per source–target pair) showing F1 for GCN, GraphSAGE, GAT under Regime A (in-domain) and Regime C (5%-fine-tune). The y-axis is F1 ranging 0.30–0.80; nine groups along the x-axis (FIN, CYB, EPI for Regime A; six transfer pairs for Regime C). Error bars indicate 95% bootstrap CIs. A horizontal dashed line at the in-domain mean of GraphSAGE on EPI marks the practical "good-enough" floor used in the discussion.]

Figure 3. Performance comparison across the nine training configurations. Error bars are 95% bootstrap confidence intervals over 5 seeds.

### 5.3. Ablation: attention heads, sampling sizes, and feature drift

We ran two ablations. First, varying GAT's first-layer head count from 1 to 8 produced a non-monotonic curve: 4 heads were optimal for FIN, 2 for CYB, and 1 for EPI. The pattern matches prior empirical evidence that attention-mechanism strategies must be tuned per task [69]. Second, varying GraphSAGE's first-layer sample size from 5 to 50 showed diminishing returns past 25 on CYB and EPI, consistent with the effective-receptive-field arguments common in the literature.

We also probed feature drift by computing MMD between source-domain and target-domain encoder activations. MMD was strongly anti-correlated with transfer F1 (Spearman  $\rho = -0.71$  across configurations), confirming that representation-level domain shift, not simply label shift, drives the transfer collapse on EPI. This finding has practical implications for domain-mixture training, which we found to be most effective when the two source domains had small pairwise MMD—FIN and CYB combined produced a stronger CYB→EPI starting point than FIN alone.

### 5.4. Robustness to label noise

We synthetically corrupted source labels with flip rates in {0%, 5%, 10%, 20%}. GAT's transfer F1 degraded most rapidly past 10%, while GraphSAGE was more robust. The pattern is reminiscent of memory-poisoning propagation results in multi-agent systems [60] in which attention-based aggregators propagate corrupted state more aggressively than sample-based ones. This is a consideration for any deployment where source-domain labels are produced by another upstream model rather than human annotators.

## 6. Discussion

### 6.1. Why GAT wins on FIN and GraphSAGE wins on EPI

The pattern in Section 5 is not just architectural taste. Financial counterparty graphs concentrate informative signal on a minority of edges—exposure to a particular distressed counterparty matters far more than exposure to a typical one. Attention is the natural inductive bias for that. Contact graphs, in contrast, distribute signal across many roughly-equivalent edges and have heavy-tailed degree distributions; sampling-based aggregation handles this gracefully without spending capacity on attention scores that all converge toward uniform. We did not test all GNN variants—graph isomorphism networks and graph transformers are obvious omissions—but the directional finding is consistent with the multi-modal-attention biomarker literature [70] and the cardiovascular-CT deep-learning translation review [71].

### 6.2. Privacy, federation, and cross-institutional sharing

In all three subtasks, raw graphs would normally be confidential. Federated learning is the obvious operational answer, with the privacy guarantees and clinical-deployment caveats discussed for medical AI [51, 49], for healthcare differential-privacy budget allocation [50], for cross-institutional financial monitoring [53], and for general practical implementation [55]. Differential privacy applied to feature attribution [72] and to creator-platform revenue transparency [73] illustrates that DP can preserve interpretability when configured carefully; the same applies to multimedia content processing in clouds [74] and to mobile advertising click prediction [75]. PII detection in clinical text [76] and privacy-aware rare-disease patient discovery [54] further inform the feature-handling pipeline. Federated transparent adaptive workflow optimizers that explicitly reduce third-party dependencies [77] suggest that intermediation can itself be a privacy attack surface, an angle worth integrating into future cross-institutional risk pipelines. For collaborative document classification under DP, gradient-clipping schedules have been shown to matter [52]; the analogous lesson for cross-domain GNNs is that gradient-clip thresholds should be tuned at the encoder level, not globally. Privacy-preserving evaluations of customer-service AI applications [78] and of multi-party financial transaction analytics [79] underscore the same point.

### 6.3. Fairness and auditability

A model that flags contagion-positive nodes can have disparate impact, particularly when nodes correspond to people (EPI) or to small businesses (FIN). Fairness-aware feature attribution methods<sup>[55]</sup> and fairness-accuracy trade-off evaluations under reweighting and resampling<sup>[56]</sup> are therefore relevant operationally. In healthcare, fairness-aware multimodal fusion for chronic-disease risk prediction<sup>[57]</sup> provides a template. We recommend that any deployment include the prevalence-parity audit reported in Section 4.4 plus an explainability layer in the spirit of feature-attribution explainability for market-risk stress scenarios<sup>[45]</sup>. Compliance-side support—automated data governance with intelligent risk reporting<sup>[80]</sup>, identifying jurisdiction clauses in cross-border financial contracts<sup>[81]</sup>, detecting implicit compliance violations across borders<sup>[82]</sup>, and contract-review IPO-audit analytics<sup>[83]</sup>—rounds out the auditability requirements typical of regulated deployment. Causal effect evaluation in welfare-program contexts<sup>[84]</sup> reminds us that a contagion model's output is itself an intervention, not just a forecast.

### 6.4. Multi-modal and multi-source feature fusion

We folded in textual signals from disclosure documents and behavioral telemetry, but did not exhaust the possible fusion sources. Multi-source data fusion approaches for credit-default early warning<sup>[7]</sup>, multi-source short-term demand forecasting<sup>[85]</sup>, multimodal cardiovascular-disease risk prediction<sup>[86]</sup>, and enhanced multi-modal feature fusion for early-stage cancer detection<sup>[87]</sup> all illustrate that fusion gains depend on careful weighting of signal reliability—exactly the attention-mechanism question Section 5.3 raised. Latency-adaptive feature-fusion weight allocation under bandwidth constraints for V2X cooperative 3D detection<sup>[88]</sup> is an instructive analogue: in cyber and financial settings, network bandwidth (or telemetry quota) genuinely constrains which features are available at inference time. Multi-sensor reliability assessment under adverse weather<sup>[89]</sup> and lightweight LiDAR-camera evaluation<sup>[90]</sup> provide further design guidance.

### 6.5. Robustness, drift, and operational monitoring

In production, contagion-detection pipelines drift over time. Empirical evaluation of multi-source monitoring signals for Kubernetes-based microservices<sup>[67]</sup> and statistical anomaly detection for payroll data migration<sup>[91]</sup> show practical patterns for operational drift detection that translate directly to risk-pipeline monitoring. Incremental feature extraction with time-decay weighting<sup>[33]</sup>, adaptive anomaly-detection thresholds for financial data quality<sup>[64]</sup>, and real-time fraud scoring on behavioral sequences<sup>[32]</sup> all share a common structure: a fast online layer that flags drift and a slower batch layer that retrains. Adaptive threshold algorithms in cardiovascular wearables<sup>[92]</sup> illustrate the same architecture in a healthcare context. For burst-load resource scheduling that affects whether the online layer can keep up, work on intelligent prediction and dynamic scheduling under burst loads<sup>[93]</sup> and on energy-efficient cloud scheduling via convex optimization<sup>[94]</sup> is directly relevant. So is server power-consumption prediction with dynamic adjustment<sup>[95]</sup>, grid-aware carbon-variable workload scheduling [96, 97], and adversarial-conditions illumination-normalization<sup>[98]</sup>. Path-optimization work for carbon-constrained last-mile delivery<sup>[99]</sup> uses the same mixed offline-online decomposition. Performance-benchmarking of depth-estimation algorithms in unstructured environments<sup>[100]</sup> demonstrates the importance of stress-testing across operating conditions.

### 6.6. Cross-domain method-borrowing

A central practical claim of this paper is that risk-domain modelers should examine adjacent literatures more carefully than they currently do. Several borrowings illustrate the pattern.

*From the LLM and NLP literature.* Retrieval-augmented generation has been shown to mitigate hallucinations in medical question answering through medical-terminology definition grounding<sup>[101]</sup>; the same idea—grounding model outputs in a vetted ontology—applies to risk-event taxonomies. Context-aware semantic ambiguity resolution in cross-cultural dialogue<sup>[102]</sup> is methodologically similar to disambiguating sectoral language in financial filings<sup>[103]</sup>. Comparative evaluation of zero-shot and few-shot LLM performance in low-resource translation<sup>[104]</sup>, prompt-specificity effects on LLM-generated code<sup>[105]</sup>, and prompting-strategy comparison for code generation<sup>[106]</sup> all bear on prompt design for risk-narrative classification. Discrete-diffusion versus autoregressive text generation<sup>[107]</sup> is relevant when generating synthetic risk scenarios for stress testing. Tenant legal inquiry classification<sup>[108]</sup> and customer-service dialogue quality assessment<sup>[109]</sup> illustrate that text classification at scale must be paired with quality-monitoring loops. NLP-driven psychological-contract risk detection in cross-cultural teams<sup>[110]</sup> connects the textual and behavioral channels.

*From the agentic-AI and multi-agent literatures.* Sequential cooperative multi-agent online learning<sup>[47]</sup>, web-agent reinforcement learning under multi-cost and failure-risk constraints<sup>[48]</sup>, continuous reorganization of agent memory under distributed change<sup>[111]</sup>, and the agentic-AI domain survey<sup>[46]</sup> together describe an emerging paradigm in which contagion-detection components are agents that share state. Memory-poisoning propagation<sup>[60]</sup> and LLM jailbreak-evolution security<sup>[112]</sup> flag that this paradigm has its own attack surface.

*From healthcare and biomedicine.* Anatomy-aware contrastive pre-training across multi-modal imaging<sup>[61]</sup> is directly transferable to graph pretraining when nodes carry partial-modality features. Clinical trial recruitment

via multi-modal screening [113] is structurally close to high-throughput risk-flagging. Bayesian optimization for nanobody screening [114] and deep-reinforcement-learning-driven efficacy-toxicity balance for personalized drug combinations [115] illustrate decision-pipeline patterns that translate to portfolio rebalancing. AI-enhanced detection of dynamic structural changes in inflammatory protein interfaces [116] and adaptive dose optimization for LED photodynamic therapy via reinforcement learning [117] use closed-loop adjustment patterns relevant to dynamic hedging. LED medical-imaging noise suppression [118] and dose optimization with uncertainty quantification in IMRT [119] further demonstrate how uncertainty-aware deep learning translates to risk-adjusted decisioning. AI-enhanced clinical-trial recruitment [113] and integrating ovarian-reserve biomarkers for gonadotoxicity [120] complete a picture in which biomedical pipelines and financial-risk pipelines share more architecture than they share vocabulary. Recurrence-time prediction with attention-enhanced LSTM for breast cancer [121] and personalized ovarian-stimulation protocol optimization [122] both embody the time-to-event modeling style that the contagion-lift metric in Section 4.4 implicitly assumes. Healthcare data-quality governance [123] and OCR-multi-engine collaboration for medical document classification [124, 125, 126, 127] provide patterns for data integrity that risk pipelines need, as does anomalous medical-insurance behavior detection [42].

*From animation, art, and creative media.* It may seem far afield, but adaptive animation generation for medical education [128] and cultural-intelligent dynamic medical animation for cross-lingual telemedicine [129] demonstrate generation pipelines whose conditional structures parallel scenario-generation in stress testing. AI-driven predictive animation state transitions for FPS games [130] and deep-learning-based prediction of animated facial expression communication effects [131] illustrate the latency-versus-quality trade-offs that real-time risk dashboards face. GAN-based keyframe interpolation for character animation [132] and AI-driven procedural animation generation for personalized medical training [133] are conceptually close to scenario interpolation in macro-stress generation. Adaptive cross-cultural medical animation [134] addresses internationalization concerns that cross-jurisdiction risk monitoring also faces. Style-genes generative AI for artwork authentication [135], CNN-based artwork-style classification [136], and blockchain-provenance verification for art counterfeit reduction [137] are loosely analogous to provenance-verification subproblems in supply-chain finance.

*From advertising and recommendation systems.* Detecting fraudulent click patterns in mobile in-app browsers [138] and privacy-preserving click-pattern anomaly detection [139] use the same behavioral-graph patterns as cyber lateral-movement detection. Feature-based bot-traffic and click-fraud detection [140] and adaptive optimization of advertising creatives [141] illustrate the operational scale at which behavioral fraud must be screened. Ride-hailing spatiotemporal preference modeling [142] and machine-learning-based investor-asset matching [143] both employ similarity-graph approaches that resemble Personalized PageRank baselines for risk-amplifier identification.

*From sustainability and infrastructure.* Carbon-emission analyses of retail distribution networks [144], building-energy-consumption prediction [145], AI-driven carbon-credit project quality assessment [146], and equity assessment of vulnerable populations in energy transition [147] illustrate how multi-objective optimization frames apply to infrastructure problems whose risk-contagion analogues we discuss above. Multi-objective particle-swarm optimization for renewable-enterprise site selection [148] uses the same Pareto-frontier framing that fairness-accuracy work uses. Efficient relational context perception for knowledge-graph completion [16] and oversampling-ensemble interactions on imbalanced tabular data [44] also support the recommendation that risk pipelines treat graph-completion and class-rebalancing as adjacent rather than separate tasks.

*From education, ASD-intervention, and dental engineering.* These are perhaps the most distant fields, but several methodological points carry over. Comparative transfer-learning evaluations in cross-context behavioral generalization [18], adaptive prompt-selection for autism skill acquisition [149], action recognition for ASD video therapy [150], adaptive difficulty-adjustment evaluation in social-skills training [151], context-aware classification of verbal operants [152], and adaptive learning-rate optimization in personalized educational interventions [153] together illustrate transfer-and-personalization protocols that map onto the encoder-fine-tuning regime used in Section 3.4. Dental-engineering work on attention-enhanced YOLO defect detection [154], machine-learning-based dental shade classification [155], NSGA-II for dental-resin 3D printing [156], biomechanical-property prediction [157], and interpretable ML for dental polymer formulation [158] illustrate the value of comparing many algorithms on a single quality task—exactly the spirit of the present benchmark. Performance evaluation of prompt generation for AI agents in programming education [159] reinforces the prompt-design points raised in Section 6.6.

*From data centers, image enhancement, and lightweight detection.* Carbon-aware spatiotemporal workload scheduling [97], deep-learning paradigms for low-light image enhancement [160], cross-modal artifact mining for deepfake detection [161], and unsupervised anomalous billing-pattern detection in healthcare payment integrity [162] together illustrate how problem framing—where to put the unsupervised layer, where the supervised one—translates across surface-different applications. Banking customer segmentation via deep embedding clustering [58] and feature-engineering for online-payment fraud detection [66] are particularly close to our risk subtasks. RPA-based audit-efficiency comparison [163] and DRL-driven cryptocurrency forecasting [164] complete the picture. Transportation route optimization for e-commerce returns [165] and AI-enhanced

what-if analysis for supply-chain digital twins <sup>[166]</sup>, plus AI-driven seasonal consumption forecasting <sup>[167]</sup> and cross-modal misinformation detection <sup>[168]</sup>, illustrate that supply-chain and information ecosystems can be modeled with the same graph-attention machinery as the contagion subtasks here.

We do not claim that all of these literatures should be read as direct predecessors. Rather, we take the empirical correspondences seriously: across very different domains, message passing on graphs of objects with heterogeneous features, with attention-weighted or sampled aggregation, is the dominant pattern, and the comparative benchmarking discipline that has matured in those literatures should inform risk-modeling practice.

## 6.7. Limitations

Three limitations are worth flagging. First, our SNAP-derived contact graph and our synthetic financial graph both rely on calibrated label-generation processes; while we matched moments and prevalence, real-world contagion may have richer structure than our generators capture. Second, we benchmarked only three GNN families; transformers on graphs and graph isomorphism networks are notable omissions. Third, the cross-domain comparison was bilateral; multi-source fine-tuning over more than two domains—closer in spirit to multi-agent online coordination <sup>[47]</sup> or to memory-structure reorganization in distributed change environments <sup>[111]</sup>—was outside our compute budget.

## 7. Conclusion

We presented a controlled comparison of GCN, GraphSAGE, and GAT on three structurally analogous risk-contagion subtasks: financial cross-market spillover, cyber lateral movement, and epidemic-style contact-graph spreading. In-domain, GAT leads on the financial graph and GraphSAGE leads on the cyber and contact graphs; the direction of the gap is consistent with each architecture's inductive bias. Cross-domain, financial and cyber graphs transfer reasonably well to each other under modest fine-tuning, while transfer involving the contact graph remains hard. Operational considerations—privacy, fairness, drift, and explainability—were treated not as afterthoughts but as constraints that determine whether the benchmark winner is actually deployable. We hope the benchmark, the fold-and-split protocol, and the cross-domain method-borrowing inventory are useful both for risk-pipeline practitioners and for graph-learning researchers looking for a defensible test bed beyond the standard citation networks.

Future work should extend the benchmark with graph transformers, explore federated cross-institutional training along the lines of [49, 53], and evaluate robustness under adversarial node injection and label poisoning—a particularly urgent direction in light of jailbreak-evolution work <sup>[112]</sup> and memory-poisoning evidence <sup>[60]</sup>.

## References

- [1]. Han, J., & Jia, R. (2026). AI-Enhanced Cross-Asset Liquidity Contagion Pathway Identification and Dynamic Hedging Strategy Optimization: Evidence from US Equity, Bond, and Derivatives Markets. *Journal of Computing Innovations and Applications*, 4(1), 89-96.
- [2]. Ren, W., Wu, X., & Li, J. (2025). AI-Driven Network Threat Behavior Pattern Recognition and Classification: An Ensemble Learning Approach with Temporal Analysis. *Journal of Advanced Computing Systems*, 5(9), 1-13.
- [3]. Wang, Y. (2025, December). Practical AI Approaches for Community Infection Early Warning: From Public Data to Actionable Insights. In *Proceedings of the 2025 6th International Conference on Computer Science and Management Technology* (pp. 1545-1552).
- [4]. Li, Y., Zhao, F., & Hu, J. (2026). Identifying Cross-Market Risk Contagion Amplifiers via Graph Attention Networks: Empirical Evidence from US Financial Stress Periods. *Journal of Computing Innovations and Applications*, 4(1), 164-175.
- [5]. Han, J. (2026). Network-Based Identification of Risk Contagion Pathways Between US Credit and Equity Markets During Stress Periods. *Journal of Advanced Computing Systems*, 6(2), 50-63.
- [6]. Wei, C., & Wu, C. (2024). Credit Risk Transmission Mechanism and Prevention Strategies in Supply Chain Finance: A Core Enterprise Perspective. *Artificial Intelligence and Machine Learning Review*, 5(2), 101-115.
- [7]. Han, J., & Cao, G. (2024). A Comparative Study of Multi-source Data Fusion Approaches for Credit Default Early Warning. *Artificial Intelligence and Machine Learning Review*, 5(1), 105-116.
- [8]. Chen, Y. (2024). Explainable Attack Path Reasoning for Industrial Control Network Security Based on Knowledge Graphs. *Journal of Computing Innovations and Applications*, 2(1), 128-139.

- [9]. Hu, J., & Long, X. (2024). Graph Learning-Based Behavioral Detection for Software Supply Chain Attacks. *Journal of Advanced Computing Systems*, 4(4), 49-60.
- [10]. Wang, Y. (2026). Accuracy Evaluation of Machine Learning-Based Hospital Resource Demand Forecasting During Infectious Disease Surges: A Comparative Analysis. *Journal of Science, Innovation & Social Impact*, 2(1), 314-327.
- [11]. Wang, Y. (2026). Explainable Risk Stratification for Polypharmacy-Related Adverse Outcomes in Community-Dwelling Elderly: A Rule-Enhanced Machine Learning Approach. *Journal of Sustainability, Policy, and Practice*, 2(2), 18-31.
- [12]. Wei, C., Ge, L., & Brooks, N. (2024). Graph-based Representation Learning for Financial Fraud and Anomaly Transaction Detection. *Journal of Computing Innovations and Applications*, 2(1), 153-164.
- [13]. Deng, M. (2025, September). Early Detection of Malicious Accounts on Social Platforms Based on Temporal Graph Feature Learning. In *Proceedings of the 2025 8th International Conference on Computer Information Science and Artificial Intelligence* (pp. 1320-1328).
- [14]. Deng, M. (2025). Graph-Based Temporal Behavior Analysis for Early Detection of Coordinated Malicious Accounts in Social Media Platforms. *Journal of Science, Innovation & Social Impact*, 1(2), 96-106.
- [15]. Cheng, Z. (2025). Graph Attention-Based Feature Selection for Multi-Omics Drug Target Prediction in Cardiovascular Diseases. *Journal of Science, Innovation & Social Impact*, 1(1), 294-306.
- [16]. Tu, W., Wan, G., Shang, Z., & Du, B. (2025). Efficient relational context perception for knowledge graph completion. *Applied Intelligence*, 55(15), 1005.
- [17]. Zhang, J. (2024). Performance Evaluation and Comparison of Machine Learning Algorithms for Anomalous Login Behavior Detection in Enterprise Networks. *Artificial Intelligence and Machine Learning Review*, 5(2), 77-90.
- [18]. Bai, Y., & Liu, M. (2026). A Comparative Evaluation of Transfer Learning Methods for Cross-Context Behavioral Generalization Assessment in Autism Spectrum Disorder Interventions. *Journal of Computing Innovations and Applications*, 4(1), 176-185.
- [19]. Li, Y., & Ling, Z. (2026). Real-Time Multi-Risk Early Warning for Community Banks: An Application of Ensemble Anomaly Detection and Explainable Artificial Intelligence. *Journal of Advanced Computing Systems*, 6(2), 15-27.
- [20]. Zhong, M. (2026). Multi-Dimensional Feature Analysis and Evaluation Methods for Anomalous Fund Flow Identification in Cross-Border Financial Transactions. *Journal of Science, Innovation & Social Impact*, 2(2), 1-13.
- [21]. Han, M., & Lai, J. (2026). Temporal Feature Engineering and Threshold Optimization for Early Warning in Healthcare Claims Anomaly Detection. *Journal of Advanced Computing Systems*, 6(4), 27-49.
- [22]. Li, Y., & Long, L. (2026). Lightweight AI-Driven Stress Testing for Small and Medium Financial Institutions: A Variational Autoencoder Approach with Extreme Value Theory for Macroeconomic Scenario Generation. *Artificial Intelligence and Machine Learning Review*, 7(1), 108-119.
- [23]. Huang, Y. (2025, August). Deep learning-enhanced dynamic margin period of risk prediction for counterparty credit risk management: A multi-modal approach integrating market sentiment analysis and real-time exposure assessment. In *Proceedings of the 2nd International Conference on Intelligent Computing and Data Analysis* (pp. 328-335).
- [24]. Huang, Y. (2024). Adaptive Importance Sampling for Jump-Diffusion CVA A Variance-Reduction Framework. *Academia Nexus Journal*, 3(3).
- [25]. Huang, Y. (2025). Enhanced Feature Engineering and Algorithm Optimization for Real-Time Detection of Synthetic Identity Fraud and Money Laundering in Financial Transactions. *Journal of Science, Innovation & Social Impact*, 1(1), 384-397.
- [26]. Ge, L. (2025). Efficiency Comparison of Automated Tools versus Traditional Methods in Anti-Money Laundering Compliance Auditing for Banking Institutions. *Journal of Science, Innovation & Social Impact*, 1(1), 265-277.

- [27]. Shi, X. (2025, August). Intelligent Credit Risk Assessment for Small and Medium Enterprises Based on Multi-dimensional Data Fusion. In Proceedings of the 2025 International Conference on Generative Artificial Intelligence for Business (pp. 186-196).
- [28]. Han, J. (2025, October). Multi-source Text Mining for Risk Signal Detection in Asset-Backed Securities Market: An NLP-driven Data Analytics Approach. In Proceedings of the 2025 International Symposium on Machine Learning and Social Computing (pp. 497-506).
- [29]. Cai, Y. (2025). NLP-Quantified ESG News Sentiment and Portfolio Outcomes Evidence from Real-Time Signals. *Annals of Applied Sciences*, 6(1).
- [30]. Cai, Y. (2025, June). NLP-Enhanced Predictive Analytics for UHNW Client Investment Behavior: A Risk-Aware Portfolio Optimization Approach in Volatile Markets. In Proceedings of the 2025 2nd International Conference on Digital Economy, Blockchain and Artificial Intelligence (pp. 185-191).
- [31]. Crawford, A., Cai, Y., & Langford, V. (2024). Machine Learning-Enhanced Dynamic Asset Allocation in Target-Date Investment Strategies for Pension Funds. *Journal of Computing Innovations and Applications*, 2(2), 122-135.
- [32]. Deng, M. (2025). Real-Time Fraud Risk Scoring through Behavioral Sequence Analysis: An Explainable Approach for online Transaction Security. *Journal of Sustainability, Policy, and Practice*, 1(4), 130-142.
- [33]. Zhong, M. (2024). Time-Decay Aware Incremental Feature Extraction for Real-Time Transaction Fraud Detection. *Artificial Intelligence and Machine Learning Review*, 5(3), 136-145.
- [34]. Wu, X., Li, J., & Ren, W. (2024). Risk Assessment Framework for Data Leakage Prevention Using Machine Learning Techniques. *Artificial Intelligence and Machine Learning Review*, 5(3), 55-66.
- [35]. Ren, W., Li, J., & Wu, X. (2024). Privacy-Preserving Data Analysis Using Federated Learning: A Practical Implementation Study. *Artificial Intelligence and Machine Learning Review*, 5(1), 40-50.
- [36]. Han, J. (2025). AI-Enhanced Cybersecurity for Financial Networks: A Federated Learning Implementation. *Journal of Science, Innovation & Social Impact*, 1(1), 241-252.
- [37]. Long, X. (2025). Research on Intelligent Firmware Vulnerability Detection and Priority Assessment Method Based on Hybrid Analysis. *Journal of Science, Innovation & Social Impact*, 1(1), 350-361.
- [38]. Long, X. (2026). Performance Evaluation of Anomaly-Based Detection Approaches for Zero-Day Attack Early Warning in Cloud Infrastructure. *Journal of Science, Innovation & Social Impact*, 2(1), 352-363.
- [39]. Long, X., Hu, J., & Ling, Z. (2026). A Comparative Analysis of Telemetry-Driven Anomaly Detection Approaches for Dual-Purpose Operational and Security Optimization in Edge Computing Infrastructure. *Journal of Computing Innovations and Applications*, 4(1), 79-88.
- [40]. Jia, R., Zhang, J., & Prescott, J. (2024). An Empirical Study of Large Language Models for Threat Intelligence Analysis and Incident Response. *Journal of Computing Innovations and Applications*, 2(1), 99-110.
- [41]. Liu, Y. (2025). Explainable Risk Stratification and Resource Coordination for Hospital Readmission Management through Integrated Prediction-Intervention-Evaluation Framework. *Journal of Science, Innovation & Social Impact*, 1(2), 107-118.
- [42]. Han, M. (2025). Intelligent Recognition of Anomalous Behaviors in Medical Insurance Through Deep Learning. *Journal of Science, Innovation & Social Impact*, 1(1), 410-426.
- [43]. Min, S., & Wei, C. (2023). Comparative Analysis of Filter-based Feature Selection Methods for High-Dimensional Data in Classification Tasks. *Journal of Advanced Computing Systems*, 3(8), 25-38.
- [44]. Wei, W., & Shang, Z. (2026). An Empirical Evaluation of Oversampling-Ensemble Interactions Under Varying Imbalance Ratios for Tabular Data Classification. *Artificial Intelligence and Machine Learning Review*, 7(2), 70-81.
- [45]. Li, Z., Huang, Y., & Montgomery, I. (2024). Feature Attribution-Based Explainability Analysis for Market Risk Stress Scenarios. *Journal of Computing Innovations and Applications*, 2(2), 136-150.

- [46]. Zhang, S., Jia, R., & Li, Z. (2024). Agentic AI Across Domains: A Comprehensive Review of Capabilities, Applications, and Future Directions. *Journal of Computing Innovations and Applications*, 2(1), 86-98.
- [47]. Yue, L., Xu, D., Qiu, D., Shi, Y., Xu, S., & Shah, M. (2025, December). Sequential Cooperative Multi-Agent Online Learning and Adaptive Coordination Control in Dynamic and Uncertain Environments. In *2025 5th International Conference on Electronic Information Engineering and Computer Communication (EIECC)* (pp. 692-697). IEEE.
- [48]. Ma, Q., Yue, L., Xu, S., Shi, Y., & Liu, H. (2026, January). Web Agent Agentic Reinforcement Learning Decision Model Under Multi-Cost and Failure Risk Constraints. In *Proceedings of the 2026 5th International Conference on Big Data, Information and Computer Network* (pp. 514-520).
- [49]. Han, M. (2025, December). Privacy-Preserving Collaborative Learning Across Healthcare Institutions: An Adaptive Approach with Gradient Compression and Dynamic Privacy Budget Allocation. In *Proceedings of the 2025 6th International Conference on Computer Science and Management Technology* (pp. 679-684).
- [50]. Shi, X. (2024). Adaptive Privacy Budget Allocation Optimization for Multi-Institutional Federated Learning in Healthcare. *Journal of Advanced Computing Systems*, 4(2), 50-61.
- [51]. Wei, C., & Guan, H. (2024). Privacy-Preserving Federated Learning in Medical AI: A Systematic Review of Techniques, Challenges, and the Clinical Deployment Gap. *Artificial Intelligence and Machine Learning Review*, 5(3), 124-135.
- [52]. Zhang, Q. (2025, December). Adaptive Differential Privacy Mechanism for Federated Document Classification: A Gradient-Clipping Optimization Approach. In *Proceedings of the 2025 6th International Conference on Computer Science and Management Technology* (pp. 672-678).
- [53]. Zhong, M. (2026). Privacy-Preserving Federated Learning for Collaborative Risk Monitoring Across Financial Institutions: Balancing Regulatory Compliance and Intelligence Sharing. *Journal of Sustainability, Policy, and Practice*, 2(2), 44-54.
- [54]. Pan, Z. (2024). Privacy-Aware AI for Rare-Disease Patient Discovery and Targeted Outreach: An Effectiveness Study. *Spectrum of Research*, 4(1).
- [55]. Zhong, M. (2025). Fairness-Aware Feature Attribution for Credit Scoring: A Causal Path Decomposition Approach. *Journal of Science, Innovation & Social Impact*, 1(1), 442-451.
- [56]. Wang, Z., & Lai, J. (2026). Fairness-Accuracy Trade-offs in AI Credit Scoring: A Comparative Evaluation of Reweighting and Resampling Strategies Under Multiple Fairness Constraints. *Journal of Computing Innovations and Applications*, 4(1), 117-126.
- [57]. Shi, X. (2026). Fairness-Aware Multimodal Fusion for Early Chronic Disease Risk Prediction: A Temporal Deep Learning Approach. *Journal of Science, Innovation & Social Impact*, 2(1), 217-231.
- [58]. Weng, H. (2025). Deep Embedding Clustering with Adaptive Feature Selection for Banking Customer Segmentation. *Spectrum of Research*, 5(2).
- [59]. Liang, D. (2026). Risk Level Classification of Contingent Liability Clauses in Financial Statement Notes Using NLP Techniques. *Artificial Intelligence and Machine Learning Review*, 7(1), 53-68.
- [60]. Liu, H., Xu, D., Ma, Q., Xu, S., & Qiu, D. (2026). Memory Poisoning Propagation and Repair Mechanism in Multi-Agent Collaborative Environments.
- [61]. Han, M. (2026). Anatomy-Aware Contrastive Pre-training: Leveraging Spatial Consistency for Label-Efficient Medical Image Diagnosis Across Multi-Modal Imaging. *Journal of Sustainability, Policy, and Practice*, 2(1), 55-70.
- [62]. Liang, D., Chen, Z., & Wei, C. (2026). Detecting Semantic Mismatches in XBRL Tag Mapping for SEC 10-K Filings: A Text Comparison and Historical Consistency Analysis. *Journal of Computing Innovations and Applications*, 4(1), 154-163.
- [63]. Liang, D. (2026). Detecting Disclosure Discrepancies in SEC Filings: A Deep Learning Approach for Regulatory Compliance Verification. *Journal of Sustainability, Policy, and Practice*, 2(1), 101-114.
- [64]. Zhong, M. (2025, September). Adaptive Anomaly Detection Threshold for Financial Data Quality Monitoring Based on Time Series Features. In *Proceedings of the 2025 International Symposium on Artificial Intelligence and Computational Social Sciences* (pp. 578-587).

- [65]. Zhong, M. (2026). Optimization of Anomaly Detection Algorithms for Consumer Credit Default Rates Based on Time-Series Feature Extraction. *Journal of Sustainability, Policy, and Practice*, 2(1), 44-54.
- [66]. Zhang, J. (2026). A Comparative Evaluation of Deep Learning and Ensemble Algorithms for Online Payment Fraud Detection. *Journal of Science, Innovation & Social Impact*, 2(1), 164-177.
- [67]. Cao, H., & Long, L. (2026). Empirical Evaluation of Multi-Source Monitoring Signal Effectiveness and Lead Time for Performance Degradation Prediction in Kubernetes-Based Microservices. *Journal of Advanced Computing Systems*, 6(4), 15-26.
- [68]. Chen, Y., & Lai, J. (2026). Multi-Metric Trustworthiness Evaluation of AI-Assisted Medical Imaging Diagnosis: Integrating Confidence Calibration and Distribution Shift Detection. *Journal of Global Engineering Review*, 4(1), 113-126.
- [69]. Wang, X., Liu, M., & Long, L. (2026). Effectiveness Evaluation of Attention Mechanism Strategies in Deep Learning-Based Single Image Super-Resolution. *Journal of Global Engineering Review*, 4(1), 89-98.
- [70]. Zhang, F., Ye, H., & Wei, C. (2024). Leveraging Multi-Modal Attention Mechanisms for Interpretable Biomarker Discovery and Early Disease Prediction. *Journal of Computing Innovations and Applications*, 2(2), 111-121.
- [71]. Zhang, F., Cheng, Z., & Holloway, V. (2024). Deep Learning in Cardiovascular CT Imaging: Evolution, Trends, and Clinical Translation from 2020 to 2025. *Journal of Computing Innovations and Applications*, 2(2), 88-99.
- [72]. Li, X. (2025). Privacy-Preserving Feature Attribution Explanations for Large-Scale Recommendation Systems: A Differential Privacy Approach. *Journal of Science, Innovation & Social Impact*, 1(1), 19-32.
- [73]. Zhang, J. (2025). Privacy-Preserving Revenue Transparency on Creator Platforms An  $\epsilon$ -Differential-Privacy Framework. *Spectrum of Research*, 5(2).
- [74]. Lei, Y. (2025). Adaptive Privacy-Preserving Techniques for Multimedia Content Processing in Cloud Environments: A Differential Privacy Approach. *Journal of Science, Innovation & Social Impact*, 1(1), 278-293.
- [75]. Lu, X. (2025). Research on Mobile Advertising Click-Through Rate Prediction Algorithm Based on Differential Privacy. *Journal of Science, Innovation & Social Impact*, 1(1), 362-371.
- [76]. Guan, H. (2025). Intelligent Detection and Protection of Personally Identifiable Information in Clinical Text: An Advanced NLP Approach with Optimized Attention Mechanisms. *Journal of Science, Innovation & Social Impact*, 1(2), 41-52.
- [77]. Wang, Z., & Kang, A. (2025). FTAFO: A Federated Transparent Adaptive Financial Optimizer for Reducing Third-Party Dependencies in Workflow Management. *Journal of Science, Innovation & Social Impact*, 1(1), 329-339.
- [78]. Zhang, Y. (2026). Evaluation of Differential Privacy and Federated Learning for AI-Driven Customer Service Applications. *Journal of Sustainability, Policy, and Practice*, 2(2), 55-66.
- [79]. Wu, Z., Zhang, Z., Zhao, Q., & Yan, L. (2025). Privacy-preserving financial transaction analytics with secure multi-party computation. Working Paper.
- [80]. Li, Y. (2026). Enhancing Financial Compliance Transparency through Automated Data Governance and Intelligent Risk Reporting. *Journal of Science, Innovation & Social Impact*, 2(1), 299-313.
- [81]. Zhang, H. (2026, January). Automated Identification of Jurisdiction Clauses in Cross-Border Financial Contracts: A Comparative Study of Rule-Based, Dictionary-Based, and Transformer-Based Approaches. In *Proceedings of the 2026 International Conference on Artificial Intelligence and Fintech* (pp. 241-248).
- [82]. Zhang, H., & Shi, W. (2026). Comparative Evaluation of Automated Detection Approaches for Identifying Implicit Compliance Violations in Cross-border Commercial Contract Clauses. *Artificial Intelligence and Machine Learning Review*, 7(2), 1-22.
- [83]. Liang, D., & Cai, C. (2025, December). Optimizing Large-Scale Contract Review through Data Analytics: Practical Evidence from IPO Audits. In *Proceedings of the 2025 6th International Conference on Computer Science and Management Technology* (pp. 242-249).

- [84]. Zhou, Y., & Long, L. (2026). Causal Effect Evaluation of Personalized Reminder Strategies on Government Welfare Program Enrollment: A Propensity Score Matching Approach. *Journal of Computing Innovations and Applications*, 4(1), 106-116.
- [85]. Wang, J. (2025). Multi-Source Data Fusion for Short-Term Demand Forecasting of Seasonal Retail Products: An Empirical Study Using Weather and Social Media Signals. *Journal of Science, Innovation & Social Impact*, 1(1), 340-349.
- [86]. Cheng, Z. (2025). AI Enabled Cardiovascular Disease Risk Prediction through Multimodal Data Fusion: A Predictive Analytics Approach. *Journal of Sustainability, Policy, and Practice*, 1(2), 98-109.
- [87]. Zhang, C. (2025). Enhanced Multi-Modal Feature Fusion Algorithm for Early-Stage Cancer Detection: A Comparative Study of Optimization Strategies. *Journal of Science, Innovation & Social Impact*, 1(1), 318-328.
- [88]. Guo, Y., & Wei, C. (2026). Latency-Adaptive Feature Fusion Weight Allocation Under Bandwidth Constraints for V2X Cooperative 3D Object Detection. *Journal of Advanced Computing Systems*, 6(3), 22-31.
- [89]. Guo, Y. (2025). Reliability Assessment and Adaptive Fusion Algorithm for Multi-Sensor Data in Autonomous Driving under Adverse Weather Conditions. *Journal of Sustainability, Policy, and Practice*, 1(4), 143-155.
- [90]. Guo, Y. (2025). Performance Evaluation of Lightweight Detection Algorithms on Compact LiDAR-Camera Configurations for Freight Transportation. *Journal of Science, Innovation & Social Impact*, 1(1), 398-409.
- [91]. Cao, H., & Shi, W. (2026). Statistical Anomaly Detection Approach for Field Mapping Validation in Enterprise Payroll Data Migration. *Journal of Computing Innovations and Applications*, 4(1), 137-153.
- [92]. Shi, W., & Cheng, Z. (2024). Enhanced Adaptive Threshold Algorithms for Real-Time Cardiovascular Risk Prediction from Wearable HRV Data. *Journal of Advanced Computing Systems*, 4(1), 46-57.
- [93]. Lei, Y. (2025, October). Intelligent Prediction and Dynamic Scheduling Optimization Strategy for Cloud Computing Resources under Burst Load Scenarios. In *Proceedings of the 2025 International Symposium on Machine Learning and Social Computing* (pp. 59-67).
- [94]. Lei, Y., & Holloway, V. (2024). Adaptive Learning-Enhanced Convex Optimization for Energy-Efficient Cloud Resource Scheduling. *Journal of Advanced Computing Systems*, 4(11), 73-85.
- [95]. Long, X. (2025, September). Machine Learning-Based Power Consumption Prediction and Dynamic Adjustment Strategies for Enterprise Servers. In *Proceedings of the 2025 8th International Conference on Computer Information Science and Artificial Intelligence* (pp. 1310-1319).
- [96]. Chen, Y., Chen, Z., & Zou, D. (2025). CarbonShift: Harnessing Grid Carbon Variability for Geo-Distributed Workload Scheduling. *Artificial Intelligence and Machine Learning Review*, 6(4), 18-31.
- [97]. Chen, Y., & Chen, Z. (2025). Multi-Objective Deep Reinforcement Learning for Carbon-Aware Spatiotemporal Workload Scheduling in Geo-Distributed Data Centers. *Journal of Advanced Computing Systems*, 5(10), 18-30.
- [98]. Li, Y. (2025, December). Comparative Analysis of Illumination Normalization Methods for Autonomous Driving Under Challenging Lighting Conditions. In *Proceedings of the 2025 6th International Conference on Computer Science and Management Technology* (pp. 633-639).
- [99]. Shi, W., & Wang, J. (2026). Intelligent Path Optimization for Carbon-Constrained Last-Mile Delivery: A Reinforcement Learning and Heuristic Approach. *Journal of Advanced Computing Systems*, 6(1), 19-31.
- [100]. Li, Y. (2026). Performance Benchmarking and Optimization Strategies for Depth Estimation Algorithms in Unstructured Environments. *Journal of Sustainability, Policy, and Practice*, 2(2), 32-43.
- [101]. Guan, H. (2025). Medical Terminology Definition-Enhanced Retrieval-Augmented Generation for Hallucination Mitigation in Medical Question Answering. *Journal of Science, Innovation & Social Impact*, 1(1), 222-240.
- [102]. Zhang D and Wang Y 2025 AI-driven quality assessment and investment risk identification for carbon credit projects in developing countries *Pinnacle Acad. Press Proc. Ser.* 3 76–92

- [103]. Zhang, D., & Ma, X. (2025). Machine Learning-Based Credit Risk Assessment for Green Bonds: Climate Factor Integration and Default Prediction Analysis. *Journal of Sustainability, Policy, and Practice*, 1(2), 121-135.
- [104]. Zhang, D., & Zheng, Q. (2025). Machine Learning-Based Building Energy Consumption Prediction and Carbon Reduction Potential Assessment in US Metropolitan Areas. *Journal of Industrial Engineering and Applied Science*, 3(5), 27-40.
- [105]. Zhang, D., & Zhang, F. (2025). AI-Assisted Identification and Equity Assessment of Vulnerable Population Impacts in US Energy Transition. *Journal of Advanced Computing Systems*, 5(7), 1-17.
- [106]. Zhang, D., & Feng, E. (2024). Quantitative Assessment of Regional Carbon Neutrality Policy Synergies Based on Deep Learning. *Journal of Advanced Computing Systems*, 4(10), 38-54.
- [107]. Guan, H. (2025). Context-Aware Semantic Ambiguity Resolution in Cross-Cultural Dialogue Understanding. *Journal of Sustainability, Policy, and Practice*, 1(2), 136-147.
- [108]. Zhang, D., & Zheng, Q. (2025). Machine Learning-Based Building Energy Consumption Prediction and Carbon Reduction Potential Assessment in US Metropolitan Areas. *Journal of Industrial Engineering and Applied Science*, 3(5), 27-40.
- [109]. Trinh, T. K., & Zhang, D. (2024). Algorithmic fairness in financial decision-making: Detection and mitigation of bias in credit scoring applications. *Journal of Advanced Computing Systems*, 4(2), 36-49.
- [110]. Dong, B., Zhang, D., & Xin, J. (2024). Deep reinforcement learning for optimizing order book imbalance-based high-frequency trading strategies. *Journal of Computing Innovations and Applications*, 2(2), 33-43.
- [111]. Zhang, H. (2026). A Comparative Study of NER Methods for Ownership Structure Extraction from M&A Due Diligence Documents. *Journal of Sustainability, Policy, and Practice*, 2(1), 71-86.
- [112]. Li, M., Wang, X., & Yu, M. (2025). Comparative Evaluation of Zero-Shot and Few-Shot Performance of Large Language Models in Low-Resource Language Machine Translation. *Journal of Global Engineering Review*, 3(2), 59-68.
- [113]. Zhang, D., & Zhang, F. (2025). AI-Assisted Identification and Equity Assessment of Vulnerable Population Impacts in US Energy Transition. *Journal of Advanced Computing Systems*, 5(7), 1-17.
- [114]. Shang Wen, Tianxing Tang. A Comparative Evaluation of URL-Sharing, Content Similarity, and Temporal Synchronicity Signals for Detecting Coordinated Inauthentic Behavior in Multilingual Political Discourse
- [115]. Yanhuan Chen, Tianxing Tang .Evaluating Prompt Engineering Strategies for Few-Shot Cyber Threat Intelligence Entity and Relation Extraction from Multi-Source Reports
- [116]. Tianxing Tang, Xuanyi Fu, Chuankai Luo. An Empirical Comparison of High-Order Feature Interaction Operators for Conversion Rate Prediction in Sparse, High-Cardinality Message-Ads Traffic : Accuracy, Efficiency, and Offline-Online Consistency
- [117]. Tianxing Tang, Mingzhuo Yu . A Comparative Evaluation of LLM-Generated Semantic Tags versus Classical Text Features (TF-IDF, LDA, BERT Embeddings) for User-Interest Enrichment in Short-Video Recommendation
- [118]. Tang, T., & Yu, M. (2024). A Comparative Empirical Study of Semantic Signal Enhancement Methods for User Interest Features in CTR Prediction: Applicability of TF-IDF Weighting, Sentence-BERT Embeddings, and LDA Topic Fusion. *Journal of Computing Innovations and Applications*, 2(1), 165-174.
- [119]. Li, M., Zhao, F., & Tang, T. (2024). How Prompt Specificity Affects Edge Case Handling in LLM-Generated Code: An Empirical Evaluation. *Artificial Intelligence and Machine Learning Review*, 5(4), 139-149.
- [120]. Zhao, F., Yu, M., & Luo, C. (2024). A Comparative Evaluation of Prompting Strategies for Code Generation with Large Language Models. *Journal of Global Engineering Review*, 2(1), 1-11.
- [121]. Xu, S., Zhao, F., & Wang, X. (2025). An Empirical Comparison of Generation Quality and Diversity Between Discrete Diffusion and Autoregressive Text Generation. *Artificial Intelligence and Machine Learning Review*, 6(2), 16-26.

- [122]. Zhang, H. (2025). Classifying Tenant Legal Inquiries: A Comparative Study of Traditional and Deep Learning Approaches. *Journal of Science, Innovation & Social Impact*, 1(1), 452-462.
- [123]. Zhang, Y. (2026). A Comparative Study of Machine Learning Methods for Automated Customer Service Dialogue Quality Assessment. *Journal of Science, Innovation & Social Impact*, 2(1), 328-338.
- [124]. Long, L., Zou, D., & Shi, W. (2026). NLP-Driven Psychological Contract Risk Detection in Cross-Cultural Teams: An XGBoost Approach with Cultural Adaptation. *Artificial Intelligence and Machine Learning Review*, 7(2), 43-53.
- [125]. Xu, S., Ma, Q., Liu, H., & Yue, L. (2026). Continuous Reorganization and Performance Preservation of Agent Memory Structure Under Distributed Change Environments.
- [126]. Shang, Z., Wei, W., & Bai, W. (2025). Evolving security in llms: A study of jailbreak attacks and defenses. arXiv preprint arXiv:2504.02080.
- [127]. Wei, C., & Pan, Z. (2026). Accelerating Clinical Trial Recruitment Through Automated Eligibility Screening with Multi-Modal Deep Learning. *Journal of Computing Innovations and Applications*, 4(1), 1-11.
- [128]. Ye, H. (2025). Bayesian Optimization-Based AI Framework for Nanobody Screening: Minimizing Experimental Failures in ELISA Detection Systems. *Journal of Sustainability, Policy, and Practice*, 1(4), 16-31.
- [129]. Ye, H. (2025). Deep Reinforcement Learning-Driven Efficacy-Toxicity Balance Optimization Strategy for Personalized Drug Combination in Cancer Patients. *Journal of Science, Innovation & Social Impact*, 1(1), 307-317.
- [130]. Ye, H. (2025, April). AI-Enhanced Detection of Dynamic Structural Changes in Inflammatory Protein Interfaces: A Case Study of CD11b/Mac-1 Interactions. In 2025 6th International Conference on Computer Engineering and Application (ICCEA) (pp. 2173-2180). IEEE.
- [131]. Dong, Z., & Jia, R. (2025). Adaptive Dose Optimization Algorithm for LED-based Photodynamic Therapy Based on Deep Reinforcement Learning. *Journal of Sustainability, Policy, and Practice*, 1(3), 144-155.
- [132]. Dong, Z., & Zhang, F. (2025). Deep Learning-Based Noise Suppression and Feature Enhancement Algorithm for LED Medical Imaging Applications. *Journal of Science, Innovation & Social Impact*, 1(1), 9-18.
- [133]. Zhang, C. (2024). Deep Learning Dose Optimization with Uncertainty Quantification for Intensity-Modulated Radiotherapy: A 3D Radiomics Approach. *Artificial Intelligence and Machine Learning Review*, 5(2), 116-129.
- [134]. Zhang, C., & Liu, M. (2026). Integrating Ovarian Reserve Biomarkers with Machine Learning for Gonadotoxicity Risk Prediction in Young Female Cancer Patients: A Scoping Review. *Journal of Computing Innovations and Applications*, 4(1), 127-136.
- [135]. Zhang, C., & Xiao, P. (2026). Optimizing Breast Cancer Recurrence Time Prediction with Attention-Enhanced LSTM Networks. *Journal of Advanced Computing Systems*, 6(1), 80-98.
- [136]. Zhang, C. (2025, October). Comparative Study of AI Algorithms in Personalized Ovarian Stimulation Protocol Optimization: Predictive Performance Analysis Based on Patient Baseline Characteristics. In Proceedings of the 4th International Conference on Artificial Intelligence and Intelligent Information Processing (pp. 654-662).
- [137]. Liu, Y. (2026). AI-Enhanced Healthcare Data Quality Governance: An Integrated Approach for Anomaly Detection and Integrity Verification. *Journal of Sustainability, Policy, and Practice*, 2(1), 215-229.
- [138]. Zhang, Q. (2026). Improving Classification Accuracy for Unstructured Medical Documents via Multi-Engine OCR and Deep Learning Collaboration. *Journal of Advanced Computing Systems*, 6(2), 1-14.
- [139]. Zhang, Q. (2025). Enhanced Feature Fusion and Transfer Learning for Multi-Format Government Document Classification. *Journal of Science, Innovation & Social Impact*, 1(1), 427-441.
- [140]. Zhang, Q. (2025). Comparative Analysis of Pre-Trained Language Models for Medical Document Classification and Priority-Based Workflow Routing. *Journal of Sustainability, Policy, and Practice*, 1(4), 205-221.

- [141]. Zhang, Q. (2026). Adaptive OCR Engine Selection and Evaluation for Multi-Format Government Document Digitization. *Artificial Intelligence and Machine Learning Review*, 7(1), 29-39.
- [142]. Wang, Z. (2024). Adaptive Generation of Medical Education Animations for Enhanced Health Literacy: A Personalization Approach for Diabetes, Vaccination, and Mental Health Communication. *Journal of Advanced Computing Systems*, 4(1), 30-45.
- [143]. Wang, Z. (2025). Cultural-Intelligent Dynamic Medical Animation Generation for Cross-Lingual Telemedicine Communication Enhancement. *Journal of Science, Innovation & Social Impact*, 1(1), 209-221.
- [144]. Wang, Z. (2025, April). DeepMotionNet: AI-Driven Predictive Animation State Transitions for Reducing Perceptual Latency in Competitive FPS Games. In *2025 6th International Conference on Computer Engineering and Application (ICCEA)* (pp. 01-08). IEEE.
- [145]. Wang, Z. (2025). Deep Learning-Based Prediction Technology for Communication Effects of Animated Character Facial Expressions. *Journal of Sustainability, Policy, and Practice*, 1(4), 105-116.
- [146]. Wang, Z., & Chu, Z. (2025). GAN-Based Intelligent Keyframe Interpolation Method for Character Animation: An Automated In-betweening Approach. *Journal of Science, Innovation & Social Impact*, 1(2), 29-40.
- [147]. Li, Z., & Wang, Z. (2024). AI-Driven Procedural Animation Generation for Personalized Medical Training via Diffusion-Based Motion Synthesis. *Artificial Intelligence and Machine Learning Review*, 5(3), 111-123.
- [148]. Li, Z., & Wang, Z. (2024). Adaptive Cross-Cultural Medical Animation: Bridging Language and Context in AI-Driven Healthcare Communication. *Artificial Intelligence and Machine Learning Review*, 5(1), 117-128.
- [149]. Li, J. (2026). Style Genes: Leveraging Generative AI for Artwork Authentication through Artistic Style Consistency Analysis. *Journal of Sustainability, Policy, and Practice*, 2(1), 87-100.
- [150]. Li, J. (2025). Enhanced CNN-based Feature Extraction and Classification for Chinese Artwork Styles. *Journal of Science, Innovation & Social Impact*, 1(2), 135-148.
- [151]. Li, J., Zhang, F., & Li, M. (2026). Comparative Effectiveness of Blockchain Provenance Verification on Counterfeit Reduction in Art Transactions: A Multi-Scenario Empirical Assessment. *Artificial Intelligence and Machine Learning Review*, 7(2), 82-92.
- [152]. Cao, H. (2024). Detecting Fraudulent Click Patterns in Mobile In-App Browsers: A Multi-dimensional Behavioral Analysis Approach. *Artificial Intelligence and Machine Learning Review*, 5(2), 130-142.
- [153]. Cao, H. (2024). Privacy-Preserving Click Pattern Anomaly Detection for Mobile In-App Browser Advertising Fraud. *Journal of Computing Innovations and Applications*, 2(2), 151-161.
- [154]. Jia, R., Lu, X., & Whitmore, S. (2024). Feature-Based Detection of Bot Traffic and Click Fraud in Mobile Advertising: A Comparative Analysis. *Journal of Computing Innovations and Applications*, 2(1), 140-152.
- [155]. Lu, X. (2025, August). Adaptive Optimization of Advertising Creative Visual Elements Based on Multi-dimensional User Behavior Data. In *Proceedings of the 2025 International Conference on Generative Artificial Intelligence for Business* (pp. 360-368).
- [156]. Shi, X. (2024). Spatiotemporal Preference Modeling for Ride-Hailing and Context-Aware Recommendations A Machine-Learning Framework. *Spectrum of Research*, 4(2).
- [157]. Wang, Z. (2025, October). Machine Learning-Driven Investor-Asset Matching Optimization in Commercial Real Estate Investment Decisions. In *Proceedings of the 2025 2nd International Conference on Digital Economy and Computer Science* (pp. 1110-1118).
- [158]. Wang, Y. (2025). Data-Driven Analysis of Transportation Route Efficiency and Carbon Emission Correlation in Retail Distribution Networks. *Journal of Science, Innovation & Social Impact*, 1(1), 253-264.
- [159]. Zhang, D., & Zheng, Q. (2025). Machine Learning-Based Building Energy Consumption Prediction and Carbon Reduction Potential Assessment in US Metropolitan Areas. *Journal of Industrial Engineering and Applied Science*, 3(5), 27-40.

- [160]. Zhang, D., & Wang, Y. (2025). AI-Driven Quality Assessment and Investment Risk Identification for Carbon Credit Projects in Developing Countries. *Pinnacle Academic Press Proceedings Series*, 3, 76-92.
- [161]. Zhang, D., & Zhang, F. (2025). AI-Assisted Identification and Equity Assessment of Vulnerable Population Impacts in US Energy Transition. *Journal of Advanced Computing Systems*, 5(7), 1-17.
- [162]. Long, L., & Hu, J. (2026). Multi-Objective Particle Swarm Optimization for Site Selection and Policy Subsidy Maximization of Foreign Renewable Energy Enterprises in the United States. *Artificial Intelligence and Machine Learning Review*, 7(2), 54-69.
- [163]. Bai, Y., & Xiao, P. (2026). Adaptive Prompt Selection and Fading Optimization for Autism Skill Acquisition: A Reinforcement Learning Approach. *Journal of Advanced Computing Systems*, 6(1), 32-44.
- [164]. Bai, Y. (2025, September). Deep Learning-based Action Recognition for Temporal Analysis and Intervention Effectiveness Assessment in Autism Spectrum Disorder Children's Video Therapy. In *Proceedings of the 2025 International Symposium on Artificial Intelligence and Computational Social Sciences* (pp. 307-314).
- [165]. Bai, Y. (2025). Effectiveness Evaluation of Adaptive Difficulty Adjustment Algorithms with Multimodal Feedback for Social Skills Training in Children with Autism Spectrum Disorder. *Journal of Sustainability, Policy, and Practice*, 1(4), 117-129.
- [166]. Bai, Y. (2026). Context-Aware Classification of Verbal Operants in Children with ASD Using Deep Learning. *Journal of Science, Innovation & Social Impact*, 2(1), 232-243.
- [167]. Shi, W., & Bai, Y. (2024). Adaptive Learning Rate Optimization for Personalized Educational Interventions in Autism Spectrum Disorder: A Multi-Objective Reinforcement Learning Approach. *Artificial Intelligence and Machine Learning Review*, 5(4), 128-138.
- [168]. Chung, P. T. (2025). Attention-Enhanced YOLO for Real-Time Defect Detection in 3D-Printed Dental Prostheses. *Journal of Science, Innovation & Social Impact*, 1(2), 119-134
- [169]. Chung, P. T. (2026). Comparative Evaluation of Machine Learning Algorithms for Spectrophotometric Dental Shade Classification. *Journal of Sustainability, Policy, and Practice*, 2(1), 204-214.
- [170]. Chung, P. T. (2026). Multi-Objective Optimization of Process Parameters for Dental Resin 3D Printing Using Improved NSGA-II Algorithm. *Journal of Science, Innovation & Social Impact*, 2(1), 276-287.
- [171]. Chung, P. T. (2025, December). Data Mining Methods for Biomechanical Property Prediction of Biomedical Materials Based on Optimized Feature Dimensionality Reduction. In *Proceedings of the 2025 6th International Conference on Computer Science and Management Technology* (pp. 174-180).
- [172]. Chung, P. T. (2025, December). Enhancing Dental Polymer Formulation through Interpretable Machine Learning: A Comparative Analysis of Feature Selection and Algorithm Performance. In *Proceedings of the 2025 6th International Conference on Computer Science and Management Technology* (pp. 234-241).
- [173]. Li, Z., & Chen, Z. (2025). Performance Evaluation of Prompt Generation Strategies for AI Agents in Online Programming Education. *Journal of Advanced Computing Systems*, 5(9), 14-27.
- [174]. Zou, D., Chen, Z., & Ling, Z. (2025). A Comparative Evaluation of Deep Learning Paradigms for Low-Light Image Enhancement: From CNNs to Diffusion Models. *Journal of Computing Innovations and Applications*, 3(2), 85-95.
- [175]. Weng, H., & Lei, Y. (2024). Cross-Modal Artifact Mining for Generalizable Deepfake Detection in the Wild. *Journal of Computing Innovations and Applications*, 2(2), 78-87.
- [176]. Shi, X., & Weng, H. (2024). Comparative Analysis of Unsupervised Learning Approaches for Anomalous Billing Pattern Detection in Healthcare Payment Integrity. *Journal of Computing Innovations and Applications*, 2(1), 111-127.
- [177]. Jiaying Li, Muyu Liu, Minhao Li. Comparative Evaluation of Ensemble Learning Algorithms for Visitor Engagement Prediction and Content Recommendation Optimization in Virtual Museum Environments. *Innovation & Social Impact*, 2(1),.

- [178]. Wang, X., Fu, X., & Zou, D. (2025). Passage, Sentence, or Proposition? An Empirical Comparison of Retrieval Granularity Effects on LLM Answer Accuracy in Retrieval-Augmented Generation. *Journal of Global Engineering Review*, 3(1), 81-90.
- [179]. Pengyuan Xiao, Xuanyi Fu. Comparative Evaluation of Post-Hoc Feature Attribution Methods on Tabular Financial Data: Faithfulness, Stability, and Computational Efficiency
- [180]. Yifei Li, Xuanyi Fu. Comparative Evaluation of Graph Neural Networks for Cross-Market Risk Contagion Path Identification in Multi-Layer Financial Networks.
- [181]. Tianxing Tang, Xuanyi Fu, Chuankai Luo. An Empirical Comparison of High-Order Feature Interaction Operators for Conversion Rate Prediction in Sparse, High-Cardinality Message-Ads Traffic : Accuracy, Efficiency, and Offline–Online Consistency.
- [182]. Xuanyi Fu, Tianxing Tang, Chuankai Luo. An Empirical Comparison of ReAct, Reflexion, Plan-and-Solve, and Tree-of-Thought Planning Strategies on Financial Question Answering and Numerical Reasoning Tasks,
- [183]. Xuanyi Fu, Danbing Zou. A Comparative Empirical Study of Over-Refusal Behavior in Closed-Source Large Language Models on Pseudo-Harmful Prompts
- [184]. Xuanyi Fu, Fanyi Zhao. An Empirical Comparison of Few-Shot Example Selection Strategies for In-Context Learning on Public Reasoning and QA Benchmarks
- [185]. Liu, H., Xu, D., Ma, Q., Xu, S., & Qiu, D. (2026). Memory Poisoning Propagation and Repair Mechanism in Multi-Agent Collaborative Environments. *Innovations and Applications*, 2(1), 140-152.
- [186]. Xu, S., Ma, Q., Liu, H., & Yue, L. (2026). Continuous Reorganization and Performance Preservation of Agent Memory Structure Under Distributed Change Environments. *Innovations and Applications*, 4(1), 127-136.
- [187]. Ge, L. (2024). Enhancing Financial Audit Efficiency Through RPA Implementation: A Comparative Analysis in Manufacturing Industry. *Journal of Computing Innovations and Applications*, 2(1), 62-73.
- [188]. Zhao, F., Zhang, M., Zhou, S., & Lou, Q. (2024). Application of deep reinforcement learning for cryptocurrency market trend forecasting and risk management.
- [189]. Xiao, P., Wang, Y., & Montgomery, I. (2024). Deep Reinforcement Learning for Route Optimization in E-commerce Return Management. *Journal of Computing Innovations and Applications*, 2(2), 100-110.
- [190]. Wang, J., & Jia, R. (2026). AI-Enhanced What-If Scenario Analysis in Supply Chain Digital Twins: A Multi-Objective Trade-Off Perspective on Cost, Resilience, and Carbon Efficiency. *Journal of Computing Innovations and Applications*, 4(1), 97-105.
- [191]. Wang, J. (2025, October). Artificial Intelligence-Driven Seasonal Consumption Forecasting and Resource Allocation Optimization in Luxury Brand Marketing. In *Proceedings of the 2025 2nd International Conference on Digital Economy and Computer Science* (pp. 1119-1127).
- [192]. Deng, M., & Zou, D. (2026). Application of Cross-Modal Content Consistency Verification in Social Media Misinformation Detection. *Artificial Intelligence and Machine Learning Review*, 7(1), 40-52.
- [193]. Deng, M., & Xu, S. (2026). Temporal-Structural Propagation Graph Analysis for Coordinated Misinformation Campaign Detection and Source Attribution in Social Networks. *Journal of Advanced Computing Systems*, 6(5), 1-11.