Review Article

Artificial Intelligence for Systemic-Risk Monitoring in the U.S. Financial System: A Narrative Review of Methods, Data Sources, and Supervisory Applications

Vera Ogette [1, ✉], Joan E. Obi [2]

1. Robinson College of Business, Georgia State University, Atlanta, USA
2. Department of Accountancy, University of Nigeria, Nsukka, Enugu, NGA

Received: May 02, 2026 | Review began: July 22, 2026 | Review ended: August 13, 2026 | Published: September 23, 2026

Abstract

Systemic-risk monitoring in the United States was reorganized after the 2007-09 financial crisis, and artificial intelligence (AI) is now becoming part of that supervisory landscape. This narrative review examines how AI and machine learning bear on U.S. systemic-risk monitoring. Sources spanned peer-reviewed methodological and empirical studies, primary U.S. regulatory and supervisory documents, and policy reports. International studies served as comparators where U.S.-specific evidence was thin. This narrative review addresses three questions: which methods and data sources support systemic-risk monitoring, under what conditions machine learning improves early-warning prediction relative to conventional statistical models, and what limits the supervisory use of AI-based monitoring. Evidence assembled here describes a developed post-crisis architecture built around financial-stability oversight, market-based systemic-risk measures, stress-testing tools, supervisory data collections, and emerging AI governance frameworks. Evidence for machine learning-based early warning is promising but uneven. Reported performance varies with jurisdiction, crisis definition, sample period, data source, and evaluation design, and many favorable findings come from cross-country or non-U.S. settings. Data infrastructure is the strongest recurring limitation. Coverage, entity identification, linkage across institutions and markets, interoperability, and reporting frequency shape what both conventional and AI-enabled monitoring systems can observe. This review contributes a U.S.-focused synthesis connecting systemic-risk measurement, supervisory data infrastructure, and AI governance, and it concludes that reliable AI-assisted monitoring depends not only on model development but also on more standardized, linkable, and timely supervisory data.

Categories: Banking and financial services, Computational Intelligence and Information Management, Risk Management

Keywords: systemic risk, financial stability, artificial intelligence, machine learning, dodd-frank, financial stability oversight council (fsoc), office of financial research (ofr), model risk management

Introduction And Background

The 2007-09 financial crisis exposed a gap in U.S. financial regulation. Oversight concentrated on the safety of individual institutions, while no single body monitored the stability of the system as a whole, and gaps between regulators left room for risk to accumulate unobserved (Financial Stability Oversight Council, 2026). Congress answered that failure with the Dodd-Frank Wall Street Reform and Consumer Protection Act, a comprehensive financial-regulatory statute enacted in 2010, whose stated purpose was to promote financial stability through improved accountability and transparency (U.S.

Congress, 2010). Two institutions were created to pursue this goal: the Financial Stability Oversight Council (FSOC), to coordinate macroprudential oversight, and the Office of Financial Research (OFR), to supply the data and analysis the Council would need (U.S. Congress, 2010).

More than a decade on, that apparatus is increasingly oriented toward artificial intelligence (AI). A marked rise in AI adoption across financial institutions has been attributed by the Council's 2025 annual report to recent technological advances, and member agencies are directed to its Artificial Intelligence Working Group (Financial Stability Oversight Council, 2025). Machine learning (ML) methods have entered the research literature on crisis prediction, and federal policy on AI in finance has accumulated across a sequence of Treasury, National Institute of Standards and Technology (NIST), and interagency frameworks. Supervisory interest now runs in two directions at once: AI is both a source of new risk requiring oversight and a tool for conducting it.

These methods have advanced faster than the supervisory data on which they depend, and the two are seldom examined together in the United States. Existing reviews tend to be organized around methods or around institutions. That leaves two practical questions unresolved: whether AI materially improves systemic-risk monitoring and what most constrains its use in supervision.

The review focuses on how systemic risk is measured, how early-warning models perform, and what limits their use in U.S. supervision. A narrative review is adopted because these questions call for integrating evidence drawn from very different sources and designs into a single interpretive account rather than pooling comparable studies. The rationale for choosing this design over systematic, scoping, and bibliometric alternatives is set out in the Methods section. International studies are used as comparators where U.S.-specific evidence is limited.

Review

Methods

This study used a narrative review design because the material relevant to systemic-risk monitoring spans empirical finance, ML studies, supervisory publications, statutes, and policy frameworks. These sources do not share a common study design or outcome that would support statistical pooling. The review was therefore designed to examine the evidence across these different literatures while keeping the U.S. supervisory setting as its primary frame.

Sources were gathered between June 1 and June 12, 2026. Google Scholar and Crossref were used to locate peer-reviewed studies, which were then retrieved from the relevant journal or publisher websites. Regulatory and policy material came from the Federal Reserve, U.S. Department of the Treasury, OFR, Securities and Exchange Commission, NIST, Financial Stability Board, International Monetary Fund, U.S. Congress, and U.S. Government Accountability Office. Reference lists of central studies were also checked for additional methodological and policy sources.

Searches used combinations of the terms “systemic risk,” “financial stability monitoring,” “macroprudential monitoring,” “early-warning models,” “banking crisis prediction,” “machine learning,” “artificial intelligence,” “supervisory data,” “Financial Stability Oversight Council,” “Office of Financial Research,” “CoVaR,” “SRISK,” “connectedness,” “financial stress index,” “model risk management,” and “AI risk management framework.” The main period of interest was 2010 through June 2026, reflecting the post-Dodd-Frank supervisory environment. Earlier studies were retained when they were needed to explain the development of established systemic-risk measures.

Sources were retained when they addressed systemic-risk measurement, financial-stability monitoring, supervisory data, banking-crisis or early-warning prediction, or the governance of AI in financial supervision. General commentary and material without a clear connection to these questions were excluded. U.S. evidence was prioritized. Studies using foreign or multinational samples were retained when they provided useful comparisons, particularly where comparable U.S. evidence was limited.

Each retained source was checked in its primary form before its findings were used. Publisher versions governed for peer-reviewed studies, and official documents governed for statutes, regulatory material, and supervisory guidance. Secondary records were used to locate sources but not as the basis for substantive claims. Where bibliographic details or reported figures differed across records, the primary source was used.

During synthesis, empirical studies were distinguished by setting. Studies using U.S. institutions, markets, or supervisory data were classified as U.S.-specific evidence. Studies based on foreign or multinational samples were treated as comparator evidence and were not taken as evidence that the same results would hold in U.S. supervisory practice. The retained material was organized around institutional architecture, systemic-risk measurement, ML early warning, supervisory data infrastructure, and AI governance. Screening counts were not recorded prospectively, so a PRISMA-style study-selection flow is not reported.

No single risk-of-bias instrument was applied across all sources because the review includes empirical prediction studies, methodological work, and regulatory and policy documents. Prediction studies were considered with attention to the definition of the predicted event, out-of-sample evaluation, model comparison, and robustness testing. Systemic-risk measurement studies were considered for empirical validation, interpretability, and relevance to the task for which they were designed to perform. Regulatory and policy claims were based on the primary official document and, where applicable, the version current at the time of the review. Findings were extracted as source-bound statements. For the revised synthesis, empirical studies were also classified by evidentiary setting and, where relevant, by the main analytical purpose of the method.

Both authors participated in source selection and interpretation. Disagreements about the relevance of a source or the interpretation of a reported finding were resolved by returning to the primary text. The Scale for the Assessment of Narrative Review Articles was used as a reporting and quality-control checklist during preparation and revision of the manuscript (Baethge et al., 2019); it was not used to assign a numerical quality score.

During preparation and revision, a large language model was used under the authors' direction for language editing, structural organization, and assistance with revised text and comparative tables. The authors reviewed the resulting material, checked substantive claims and numerical results against the cited sources, and take responsibility for the final manuscript. The AI tool was not treated as a source of evidence and is not credited as an author.

The institutional architecture of systemic-risk monitoring

A single diagnosis shaped the institutional architecture for systemic-risk monitoring in the United States: that the financial system can fail in ways that no individual-institution supervisor is positioned to see. Before the Dodd-Frank Act, enacted as Public Law 111-203 in July 2010, oversight concentrated on individual institutions and markets. No single regulator held responsibility for broader threats to financial stability, and the resulting supervisory gaps left room for regulatory arbitrage, the practice of routing activity to the most lightly regulated part of the system (U.S. Congress, 2010) (Financial Stability Oversight Council, 2026). Those vulnerabilities were subsequently exposed across the regulatory framework by the 2007-09 crisis (Financial Stability Oversight Council, 2026). This recognition built on a broader research program in financial-stability monitoring, which organizes the supervisory task around the systematic identification of vulnerabilities in the financial system (Adrian et al., 2015). The Federal Reserve's financial-stability framework similarly distinguishes shocks from vulnerabilities and organizes monitoring around asset valuations, borrowing by businesses and households, financial-sector leverage, and funding risks (Board of Governors of the Federal Reserve System, 2025).

Congress framed the Act's purpose as promoting financial stability through improved accountability and transparency (U.S. Congress, 2010), and its structural response followed from the diagnosis. Title I established the FSOC as a macroprudential coordinator charged with responding to emerging threats to the stability of the financial system, and it created the OFR to provide the data, standards, and analysis the Council and its member agencies would require (U.S. Congress, 2010) (Office of Financial Research, 2026a) (Financial Stability Oversight Council, 2025). Saguato describes the Council in similar terms, as a multiagency body created to mitigate systemic risk through coordination and envisioned as a macroprudential stabilizer for the financial system (Saguato, 2022).

This coordinating architecture responds to the fragmented structure of U.S. financial supervision. Banking, securities, derivatives, and insurance activities are overseen by different federal and state authorities, making coordination across regulators central to system-wide oversight. FSOC brings these authorities together to identify emerging threats, share information, and coordinate responses to risks that may extend across individual regulatory jurisdictions (Financial Stability Oversight Council, 2025) (U.S. Government Accountability Office, 2023). Macroprudential oversight, in this setting, refers to supervision of risks that can threaten the financial system as a whole rather than only the safety of an individual institution. Figure 1 places these bodies at the top of the architecture traced in this review, above the measurement methods they use and the data infrastructure beneath them.

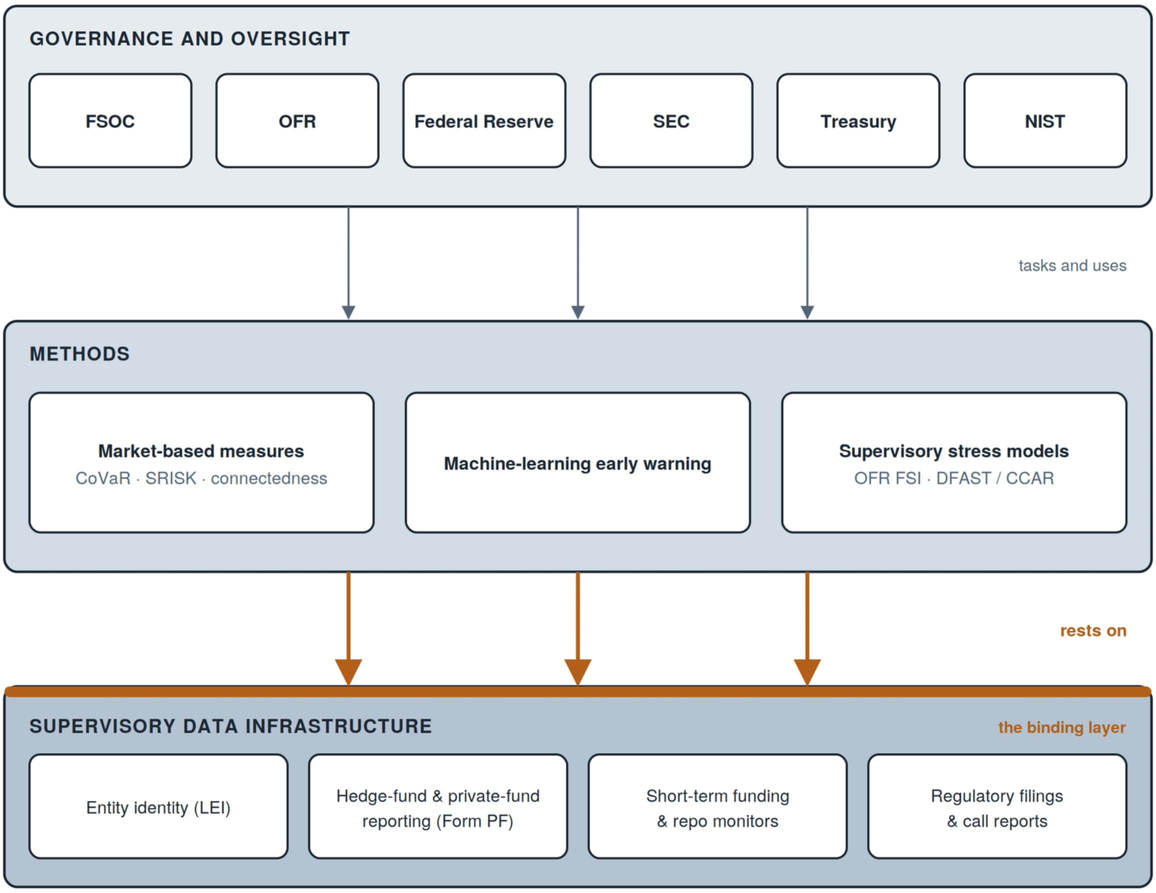

FIGURE 1: The U.S. systemic-risk monitoring architecture

Source: Authors' creation based on reviewed literature

CCAR, Comprehensive Capital Analysis and Review; CoVaR, Conditional Value at Risk; DFAST, Dodd-Frank Act Stress Test; FSI, Financial Stress Index; FSOC, Financial Stability Oversight Council; NIST, National Institute of Standards and Technology; OFR, Office of Financial Research; SEC, Securities and Exchange Commission; SRISK, Systemic Risk

Academic measurement methods for systemic risk

Institutional reforms explain who was given responsibility for monitoring systemic risk. They do not, by themselves, resolve how it should be measured. That question has largely been taken up in the academic literature, where influential approaches infer system-wide vulnerability from market prices, balance-sheet conditions, and patterns of connectedness among financial institutions.

How to cite this article:
Ogette V, Obi J E (September 23, 2026) Artificial Intelligence for Systemic-Risk Monitoring in the U.S. Financial System: A Narrative Review of Methods, Data Sources, and Supervisory Applications. Cureus J Bus Econ 3 : es44404-026-00293-9. DOI https://doi.org/10.7759/s44404-026-00293-9

These approaches do not address a single analytical task. Market-based measures such as Conditional Value at Risk (CoVaR) and the systemic risk conditional capital-shortfall measure primarily quantify or rank an institution's contribution to system-wide vulnerability, whereas connectedness measures examine how shocks propagate among institutions or sectors. Other approaches are designed for forecasting, including early-warning models that estimate the likelihood of future banking crises or macrofinancial stress. Still others evaluate existing measures, synthesize methodological alternatives, or examine how analytical tools might support supervisory decisions. Table *1* therefore summarizes the reviewed methods by primary analytical purpose, evidentiary setting, and principal finding.

How to cite this article:
Ogette V, Obi J E (September 23, 2026) Artificial Intelligence for Systemic-Risk Monitoring in the U.S. Financial System: A Narrative Review of Methods, Data Sources, and Supervisory Applications. Cureus J Bus Econ 3 : es44404-026-00293-9. DOI https://doi.org/10.7759/s44404-026-00293-9

References	Measure/method	Primary analytical purpose	Evidence setting/role	Key finding
(Adrian and Brunnermeier, 2016)	CoVaR; market-price conditional VaR	Quantify and rank an institution's contribution to systemic risk	U.S.-specific empirical evidence – U.S. financial firms	The 2006 IV-based reading would have anticipated more than one-third of realized CoVaR during the 2007-09 crisis
(Brownlees and Engle, 2017)	SRISK; conditional capital shortfall	Estimate and rank institutions by expected capital shortfall under systemic stress	U.S.-specific empirical evidence – U.S. market data	Identified Fannie Mae, Freddie Mac, Morgan Stanley, Bear Stearns, and Lehman as leading contributors by 2005 Q1; aggregate SRISK leads deterioration in real activity
(Acharya et al., 2017)	Systemic expected shortfall (SES); theoretical model	Formalize an institution's expected contribution to system-wide undercapitalization	Conceptual/theoretical evidence	Defines each firm's systemic contribution through expected undercapitalization when the financial system is undercapitalized
(Billio et al., 2012)	PCA and Granger-causality connectedness	Identify interconnectedness and channels of shock transmission	U.S.-specific empirical evidence – four U.S. financial sectors	Financial sectors became highly interrelated before the crisis; banks were disproportionate transmitters of shocks
(Diebold and Yılmaz, 2014)	Variance-decomposition connectedness	Measure the direction and magnitude of risk spillovers among institutions	U.S.-specific empirical evidence – U.S. financial institutions	Directed, weighted connectedness networks were associated with measures of system stress
(Bisias et al., 2012)	Survey of systemic-risk measures	Catalogue and compare alternative approaches to systemic-risk measurement	Methodological synthesis	Catalogs 31 quantitative systemic-risk measures spanning differing definitions of systemic risk

(Benoit et al., 2017)	Critical taxonomy of systemic-risk measures	Classify measurement approaches and identify conceptual and data limitations	Methodological synthesis	Identifies the gap between isolating individual sources of risk and aggregating them; links progress to improved data disclosure
(Acharya et al., 2025)	Empirical test of market-based systemic-risk measures	Assess whether systemic-risk measures are better suited to ranking institutions or forecasting stress	U.S.-specific empirical evidence – U.S. firms and stress episodes, 1895–2023	Measures perform strongly for cross-sectional ranking during stress but are weaker as standalone forecasts of whether stress will occur
(Giglio et al., 2016)	Dimension-reduction indexes from systemic-risk measures	Forecast downside macroeconomic outcomes using aggregated systemic-risk information	Mixed U.S./international evidence – U.S. and European data	Few individual measures forecast lower-tail outcomes; compressed indexes forecast 10th-percentile industrial-production growth shocks with out-of-sample R^2 of roughly 5-20%
(Kou et al., 2019)	Survey of ML approaches to systemic risk	Synthesize big-data, network, and sentiment methods for systemic-risk and contagion analysis	Methodological synthesis	Maps ML, big-data, network, and sentiment approaches relevant to systemic-risk analysis and market regulation
(Nyman et al., 2021)	Algorithmic text and sentiment analysis	Detect buildup and reversal of pre-crisis market sentiment	International comparator – United Kingdom	Changes in emotional content trace the formation and collapse of pre-crisis exuberance
(Hanley and Hoberg, 2019)	Computational linguistics applied to regulatory filings	Detect emerging financial-sector risk and predict institution-level outcomes	U.S.-specific empirical evidence – U.S. banks and regulatory filings	An interpretable filing-based measure detected emerging risk by the mid-2000s and predicted returns, bank failures, and volatility
(Bluwstein et al., 2023)	ML early-warning models; extreme trees and other nonlinear methods	Forecast systemic banking crises and compare ML with conventional logit models	Cross-country comparator – 17 developed countries	Most nonlinear ML models outperform logit; credit growth and yield-curve slope lead prediction.

(Holopainen and Sarlin, 2017)	Early-warning model comparison and ensembles	Compare conventional and ML methods for banking-crisis early warning	International comparator – European sample	ML methods, including k-nearest neighbors and neural networks, and especially ensembles, outperform conventional approaches
(Wang et al., 2024) (Tang et al., 2024)	Interpretable ML and connectedness features	Improve systemic-risk prediction using nonlinear models and network information	International comparator – China	Report gains in systemic-risk prediction using national samples whose institutional settings differ from that of the United States
(Beutel et al., 2019)	Logit versus four ML methods	Benchmark ML against conventional logit for systemic banking-crisis forecasting	Cross-country comparator – 15 advanced economies	Logit is difficult to outperform in recursive out-of-sample evaluation; the result remains robust across multiple specification choices
(Buckmann et al., 2021)	Conceptual comparison of human and AI	Examine human-AI complementarity in financial-stability analysis and decision-making	Conceptual/theoretical evidence	Frames the problem as designing effective human-AI partnerships rather than identifying a single superior approach
(Daníelsson et al., 2022)	Conceptual analysis of AI and financial stability	Assess AI both as a systemic-risk source and as a tool for financial-stability authorities	Conceptual/theoretical evidence	AI may amplify systemic risk through procyclicality and shared models while simultaneously improving authorities' analytical capabilities

TABLE 1: Empirical and methodological studies of systemic-risk measurement and prediction

Source: Authors' synthesis of reviewed studies.
One row per study; jurisdiction and sample shown where applicable.
AI, Artificial Intelligence; CoVaR, Conditional Value at Risk; ML, Machine Learning; PCA, Principal Component Analysis; SRISK, Systemic Risk; VaR, Value at Risk

Adrian and Brunnermeier's CoVaR captures the change in the value at risk of the financial system conditional on an individual institution being under distress, and the same study reports that institution characteristics, including leverage, size, maturity mismatch, and exposure to asset-price booms, significantly predict an institution's later contribution to system risk (Adrian and Brunnermeier, 2016). Its 2006 reading would have anticipated more than a third of the realized CoVaR observed during the 2007-09 crisis (Adrian and Brunnermeier, 2016). Brownlees and Engle's systemic risk takes a complementary route, estimating the capital shortfall a firm would face conditional on a severe market decline as a function of its size, leverage, and risk; that measure flagged Fannie Mae, Freddie Mac, Morgan Stanley, Bear Stearns, and Lehman Brothers as leading contributors as early as the first quarter of 2005, well ahead of their failures (Brownlees and Engle, 2017). Acharya et al. provide the theoretical basis for this group of market-based measures, modeling how undercapitalization of the financial sector as a whole imposes costs on the real economy and defining each institution's systemic expected shortfall as its propensity to be undercapitalized when the system is undercapitalized (Acharya et al., 2017). Aggregate systemic risk, for its part, has been shown to lead to deterioration in indicators of real activity, extending the measures' reach from the relative ranking of institutions toward macroeconomic early warning (Brownlees and Engle, 2017).

A second group of measures does not rely on balance-sheet conditioning and instead draws on the statistical structure of asset returns. Billio et al. build connectedness measures from principal-components analysis and Granger-causality networks across hedge funds, banks, broker-dealers, and insurers and find that all four sectors grew highly interrelated over the decade preceding the crisis in a way that plausibly raised systemic risk (Billio et al., 2012). Diebold and Yılmaz derive connectedness instead from variance decompositions, which define weighted and directed networks among institutions and link the resulting indices to established measures of system stress (Diebold and Yılmaz, 2014). Beyond a single risk statistic, the network approach contributes directionality: Billio et al.'s measures locate banks as disproportionate transmitters of shocks to the other three sectors, a structural asymmetry that a purely institution-level score would miss (Billio et al., 2012).

Proliferation of measures eventually became a subject of study in its own right. Bisias et al. catalog 31 quantitative measures of systemic risk and observe that the catalog spans incompatible definitions of systemic risk itself (Bisias et al., 2012). Benoit et al. sharpen the point into a diagnosis, identifying a gap between approaches that study individual sources of risk in isolation and those that aggregate them, and arguing that closing it will require both encompassing theoretical models and improved data disclosure (Benoit et al., 2017). That second requirement, improved data disclosure, marks the first appearance in the measurement literature of the constraint that this review treats as central. An unresolved question underlies the catalogue: whether these measures forecast crises or merely rank institutions once stress has begun. Recent evidence is divided. Acharya et al., testing market-based measures on U.S. financial firms' stock-return data across stress episodes dated from 1895 to 2023, find the measures effective at ranking institutions relative to one another conditional on a stress episode but weaker as standalone forecasts of whether stress will arrive (Acharya et al., 2025). Giglio et al. reach a more favorable verdict on prediction, though a carefully bounded one. Across 19 candidate measures, few individual measures forecast the lower tail of macroeconomic shocks out of sample; dimension-reduction estimators that compress the cross-section into a small number of indexes do, forecasting the 10th percentile of out-of-sample industrial-production growth shocks with R^2 values between roughly 5 and 20 percent (Giglio et al., 2016). These two findings are more consistent than they first appear. Acharya et al. report that individual measures rank better than they forecast; Giglio et al. report that individual measures forecast weakly until aggregated (Giglio et al., 2016). The operative distinction lies less between ranking and forecasting than between individual measures and aggregated indexes. How much predictive weight a supervisor should place on these readings before stress is already underway remains an open question, one that recurs in sharper form once ML is applied to the same problem.

ML and AI extensions

ML has been used in this literature for several different tasks, including crisis prediction, text analysis, sentiment measurement, and network-based risk assessment. Kou et al. review these applications and their use in studying the emergence and transmission of systemic risk (Kou et al., 2019).

Nyman et al. use financial-market commentary from Bank of England sources, broker research, and Reuters news to examine changes in sentiment before periods of stress. Their Reuters sample covered January 1996 to September 2014 and averaged 6,123 articles per month. Sentiment remained unusually positive through much of the mid-2000s before deteriorating from 2007, while their measure of narrative consensus also rose ahead of the global financial crisis (Nyman et al., 2021).

Hanley and Hoberg examine U.S. bank regulatory filings rather than news or market commentary. Their emerging-risk index rose in 2005 and peaked in late 2006, before the sharp increase in market volatility seen during the crisis itself. The filings showed increasing exposure to risks involving real estate, commercial paper, credit cards, and other areas before the crisis. Banks with greater exposure to these emerging risks subsequently experienced poorer stock returns, higher failure rates, and greater volatility, with predictive associations extending as far as 36 months (Hanley and Hoberg, 2019).

Bluwstein et al. examine banking-crisis prediction using data from 17 developed countries over 1870-2016. Most of the nonlinear ML models in their comparison outperform logistic regression in out-of-sample prediction and forecasting. Credit growth and the slope of the yield curve are the most important predictors identified in their interpretation of the models (Bluwstein et al., 2023).

Holopainen and Sarlin report a similar advantage for ML using quarterly data from 15 European Union countries between 1976Q1 and 2014Q3. Their sample contains 15 systemic banking crises and uses a warning horizon of five to 12 quarters. They compare conventional statistical models with methods including k-nearest neighbors and neural networks and also test ensemble models. ML methods generally perform better on their usefulness and area-under-the-curve measures, with ensemble models performing particularly well (Holopainen and Sarlin, 2017).

Beutel et al. reach a different result. Their study covers 15 advanced economies from 1970 to 2016 and evaluates forecasts recursively from 2005Q3 to 2016Q4. The benchmark logit model is difficult to outperform across the specifications they test. Some ML models fit the estimation data extremely well but lose much of that advantage out of sample, which the authors identify as evidence of overfitting (Beutel et al., 2019).

The disagreement is partly methodological. In the Holopainen-Sarlin recursive exercise, more than 95% of the pre-crisis observations fall in the out-of-sample period and are therefore unavailable when the first models are estimated. Beutel et al. report that nearly 60% of the pre-crisis observations in their broader sample are already available at the start of their recursive exercise (Beutel et al., 2019). The studies also differ in crisis definition, validation procedure, prediction horizon, performance measures, and treatment of the relatively small number of crisis observations. Table *2* sets out those differences.

Design feature	(Holopainen and Sarlin, 2017)	(Beutel et al., 2019)	(Bluwstein et al., 2023)
Sample characteristics	Quarterly unbalanced panel; 15 EU countries; 1976Q1–2014Q3; 15 crises	15 advanced economies; 1970–2016	Annual; 17 countries; 1870–2016
Crisis definition/target	Systemic banking crises	Systemic banking crises with partly domestic origins	Banking-sector financial crises; Macrohistory Database
Prediction horizon	5–12 quarters	5–12 quarters	1–2 years
Input variables	Asset prices, leverage, macro, external and banking indicators	10 indicators covering credit/assets, macro and external factors	10 domestic variables plus global credit and yield-curve slope
Validation procedure	10-fold CV; recursive real-time test	Recursive OOS; CV for tuning	Repeated 5-fold CV; recursive forecasting
Performance metrics	Usefulness, AUC, precision/recall, accuracy, error rates	Usefulness, F-measure, AUC, Brier score	AUC, hit rate, false-alarm rate
Class-imbalance handling/evaluation	Usefulness weights crisis/non-crisis errors	Usefulness weights missed crises/false alarms	Up/downsampling in recursive forecasts
Main comparative result	ML and ensembles generally outperform conventional methods	Logit generally outperforms ML out of sample	ML generally outperforms logit

TABLE 2: Evaluation designs of key ML early-warning studies

Source: Authors' synthesis of reviewed studies

AUC, Area Under the Curve; CV, Cross-Validation; ML, Machine Learning; OOS, Out-of-Sample

Supervisory Data Infrastructure

Both the measurement and the ML literatures returned repeatedly to the same precondition: the data available to observe the system. That precondition has a concrete history. When Lehman Brothers failed in September 2008, firms could not readily assess their total exposure to it, because no industry-wide system existed to identify a single counterparty across its many legal entities (Office of Financial Research, 2021). A structural answer followed in the Legal Entity Identifier (LEI), a standardized global code that uniquely identifies a legal entity in financial transactions; since the system's launch in 2014, more than 1.9 million identifiers have been issued to entities in over 225 countries as of August 2021, giving regulators and firms a common key for tracing who is connected to whom (Office of Financial Research, 2021).

How to cite this article:
Ogette V, Obi J E (September 23, 2026) Artificial Intelligence for Systemic-Risk Monitoring in the U.S. Financial System: A Narrative Review of Methods, Data Sources, and Supervisory Applications. Cureus J Bus Econ 3 : es44404-026-00293-9.
DOI https://doi.org/10.7759/s44404-026-00293-9

On the strength of such standards, the OFR has built data collections aimed at the parts of the system most prone to runs. Its Short-term Funding Monitor organizes short-term funding-market data, the markets in which institutions borrow and lend over very short horizons, into four categories and publishes OFR estimates of activity in the tri-party and centrally cleared repo markets, in which securities are exchanged for cash under agreements to repurchase (Office of Financial Research, 2025). Unveiled in July 2024, the Hedge Fund Monitor aggregates exposures from four sources, among them SEC Form PF filings, the confidential reports private funds file on their risk exposures, and Commodity Futures Trading Commission futures data, and concentrates on qualifying hedge funds, defined as those holding at least $500 million in net assets under a large adviser (Office of Financial Research, 2024). Form PF itself, adopted in 2011 to give the Council and the SEC visibility into private-fund risk, was amended in 2023 to tighten reporting for large hedge fund advisers with at least $1.5 billion in hedge fund assets (U.S. Securities and Exchange Commission, 2023). For large banks, the OFR's Bank Systemic Risk Monitor complements these collections by bringing together systemic-importance scores, the OFR Contagion Index, leverage measures, and short-term wholesale-funding indicators in a single monitoring tool (Office of Financial Research, 2026b).

Hedge funds and other non-bank financial institutions can transmit stress outside the traditional banking perimeter, and the private-fund reporting regime exists to make that activity visible to supervisors. The Securities and Exchange Commission oversees securities markets and investment advisers, while the Commodity Futures Trading Commission collects data on futures and derivatives markets; together, these sources extend systemic-risk monitoring beyond banks to the entities and markets where stress increasingly originates.

Beyond storing data, these collections also feed supervisory models. Drawing on 33 financial-market variables, including yield spreads and valuation measures, the OFR's Financial Stress Index uses a logistic-regression framework calibrated against government-intervention dates over 2000-18 to convert them into a continuous read on system stress (Monin, 2019). At the center of the supervisory toolkit sits the stress-testing apparatus. The Dodd-Frank Act Stress Test is a forward-looking quantitative exercise used to assess how adverse economic conditions would affect large banking organizations, while the Comprehensive Capital Analysis and Review evaluates whether the largest holding companies have adequate capital-planning processes and sufficient capital to withstand stress (Board of Governors of the Federal Reserve System, 2019). Both function as supervisory tools that test the resilience of major banks under hypothetical adverse scenarios, rather than as market forecasts.

Together, these collections show that the hardest problem in systemic-risk monitoring is increasingly the data itself. Flood et al. survey the data-management challenges facing macroprudential supervisors and reach a pointed conclusion: as oversight broadens from individual institutions to the system as a whole, the data problem grows in both dimensions at once, quantitatively in the volume of information required and qualitatively in the difficulty of integrating and standardizing it (Flood et al., 2013). A more recent strand names the implication directly. Data-centric AI, as characterized by Jakubik et al., treats the systematic improvement of data at scale, in place of further refinement of model architecture, as the route to effective systems (Jakubik et al., 2024). Supervisory data infrastructure is thus a first-order condition for both conventional and AI-based monitoring. The volume of available information matters less than whether observations can be linked across entities, markets, and reporting systems with sufficient timeliness for supervisory use.

Federal AI Policy and Governance

If data infrastructure determines what AI-based monitoring can observe, governance determines how such systems should be developed, validated, and controlled. Federal policy responses to AI in finance have become an important part of the monitoring architecture, even where they do not directly solve the underlying data problem. Treasury has issued two reports in close succession: a March 2024 assessment of AI-related cybersecurity and fraud risks, prepared in response to Executive Order 14110 and grounded in 42 industry interviews, and a December 2024 report synthesizing 103 comment letters on the uses, opportunities, and risks of AI across the sector (U.S. Department of the Treasury, 2024a) (U.S. Department of the Treasury, 2024b). That second report stresses that generative AI broadens opportunities while amplifying risks tied to data privacy, bias, and third-party providers and builds explicitly on the first (U.S. Department of the Treasury, 2024b). Standard-setting has moved in parallel. The NIST AI Risk Management Framework offers a voluntary

How to cite this article:
Ogette V, Obi J E (September 23, 2026) Artificial Intelligence for Systemic-Risk Monitoring in the U.S. Financial System: A Narrative Review of Methods, Data Sources, and Supervisory Applications. Cureus J Bus Econ 3 : es44404-026-00293-9.
DOI https://doi.org/10.7759/s44404-026-00293-9

structure for incorporating trustworthiness into the design, development, use, and evaluation of AI systems, organized around four functions: Govern, Map, Measure, and Manage (Tabassi, 2023). In February 2026, that framework gained a sector-specific adaptation when Treasury, through the interagency AI Executive Oversight Group, released a Financial Services AI Risk Management Framework mapping the NIST functions onto the operational, regulatory, and consumer-protection realities of finance (U.S. Department of the Treasury, 2026).

This governance framework did not develop in isolation. U.S. banking supervision is shared across several agencies: the Federal Reserve supervises bank holding companies and other large financial institutions, the Office of the Comptroller of the Currency (OCC) supervises national banks, and the Federal Deposit Insurance Corporation (FDIC) supervises insured depository institutions and administers deposit insurance. Model-risk discipline within this structure carries a fifteen-year supervisory history. In 2011, the Federal Reserve and the OCC issued joint Supervisory Guidance on Model Risk Management, setting expectations for model development, validation, and governance that have applied to the statistical and ML models used in supervision (Board of Governors of the Federal Reserve System, 2011). That guidance was revised in April 2026, when the Federal Reserve, the OCC, and the FDIC jointly issued SR 26-2, superseding the 2011 guidance and emphasizing a risk-based approach tailored to a banking organization's model-risk profile and the size and complexity of its operations (Board of Governors of the Federal Reserve System, 2026). The current baseline for supervising model risk is tailored and principles-based, and these AI frameworks extend an established supervisory approach rather than creating one for the first time.

International bodies have reached parallel conclusions from a wider vantage. The International Monetary Fund's October 2024 Global Financial Stability Report assesses how advances in AI and generative AI bear on capital-market activities (International Monetary Fund, 2024). Reviewing the same terrain, the Financial Stability Board reports that rapid AI adoption may amplify vulnerabilities, including third-party dependencies, market correlations, cyber risk, and model risk, with the potential to increase systemic risk, and calls on authorities to enhance their monitoring of AI developments and to strengthen supervisory capacity, including through AI-powered tools (Financial Stability Board, 2024). These assessments are read here as comparators, describing a global policy direction against which the U.S. trajectory can be located.

Academic work has begun to convert diagnosis into prescription. Danίelsson and Uthemann, working from a game-theoretic model, argue that AI amplifies existing vulnerabilities, among them leverage, liquidity stress, and opacity, through superior information processing, common data, speed, and strategic complementarities, and propose a concrete supervisory response: that authorities develop their own AI systems and expertise, establish direct AI-to-AI communication, implement automated crisis facilities, and monitor AI use across the system (Danίelsson and Uthemann, 2025). Aldasoro et al. take a broader view, assessing the implications of AI advances for financial stability and prudential policy and proposing a framework for upgrading financial regulation built on established principles of AI governance (Aldasoro et al., 2025).

Discussion

The Data-Infrastructure Constraint

Systemic-risk monitoring is constrained less by any shortage of models than by the condition of the data environment in which they operate. Several studies make this point. Benoit et al. locate future progress in better data disclosure (Benoit et al., 2017); Hanley and Hoberg show that a U.S. filing-based measure can generate early-warning information from regulatory text (Hanley and Hoberg, 2019); and Flood et al. describe the macroprudential data problem as both quantitative and qualitative, involving the amount of information required and the difficulty of reconciling sources that were not built to align (Flood et al., 2013). This agreement supports a data-centric interpretation of AI-based monitoring because improvements in data quality, linkage, and timeliness may matter as much as further refinement of model architecture. The pattern is one of diminishing returns. As modeling sophistication rises, monitoring capability increases but approaches an upper bound determined by the coverage, identity resolution, and frequency of the data available to supply it.

Empirical evidence does not support a general claim that ML is superior to conventional early-warning models. Bluwstein et al. (Bluwstein et al., 2023) and Holopainen and Sarlin report better performance for ML approaches, whereas Beutel et al. find logit difficult to outperform in recursive out-of-sample forecasting (Beutel et al., 2019). They note that the studies also differ in how much pre-crisis information is available when recursive forecasting begins. Table *2* shows further differences in crisis definition, prediction horizon, validation procedure, and treatment of rare crisis observations. These are not equivalent forecasting tests, so the reported performance differences cannot be attributed to the algorithms alone.

That qualification matters particularly for U.S. supervision. Much of the favorable early-warning evidence comes from European, Chinese, or multinational samples, while directly comparable U.S. evidence remains limited. Hanley and Hoberg provide one of the strongest domestic examples, showing that information extracted from U.S. bank regulatory filings can identify emerging risk and predict institution-level outcomes (Hanley and Hoberg, 2019). International studies demonstrate that ML gains are possible under some conditions, but they do not establish that those gains would persist under U.S. institutions, supervisory reporting systems, or data constraints.

These methods should be judged against the job they are meant to do. CoVaR and systemic risk rank institutions, connectedness measures trace spillovers, stress tests assess resilience under assumed shocks, and early-warning models estimate future crisis risk. Their performance also depends on the supervisory data available to them. A method that works well for one task or dataset should not automatically be treated as the best tool for another.

Implications and limitations

For supervisors, better data and stronger model governance need to develop together. Wider entity coverage, more reliable counterparty identification, improved linkage across reporting systems, and faster reporting would benefit conventional and ML methods alike. AI governance frameworks can guide model development, validation, and control, but they cannot make up for missing or poorly connected supervisory data. Improving the models without improving the underlying data is unlikely to produce the same gains (Daníelsson and Uthemann, 2025) (Flood et al., 2013) (Tabassi, 2023) (U.S. Department of the Treasury, 2026).

More U.S.-specific early-warning studies are needed, particularly studies using transparent out-of-sample evaluation. Comparisons would also be easier to interpret if studies used clearer and more consistent definitions of crises and prediction horizons and reported their validation procedures and treatment of rare crisis observations in enough detail to allow replication.

This review has the limits of a narrative design. It was not preregistered, and screening counts were not maintained prospectively, so a PRISMA-style study-selection flow cannot be reported reliably. Source selection involved author judgment, and relevant studies may have been missed. The material reviewed also spans empirical prediction studies, methodological work, statutes, supervisory publications, and policy frameworks, so no single risk-of-bias instrument was applied across all sources. Empirical studies were assessed through their methods, validation, and reported results, while regulatory and policy claims were checked against primary official documents.

Many of the studies reporting favorable results for ML early warning use non-U.S. or multinational samples. Their findings help show what these methods can achieve, but they do not establish comparable performance in U.S. supervisory settings. The discussion of AI policy is likewise limited to the guidance and frameworks available when the review was completed.

Conclusions

The United States does not rely on one method to monitor systemic risk. Supervisors use market-based measures, connectedness analysis, stress tests, and regulatory data, while AI-based methods are still being tested. The studies reviewed here do not show that ML consistently outperforms conventional early-warning models. Results vary across studies, and many of the strongest positive findings come from outside the United States.

A more consistent problem is the data. Gaps in coverage, weak entity matching, poor linkage across reporting systems, and reporting delays limit what any model can detect. Better U.S.-specific testing is still needed, but better algorithms alone will not solve the monitoring problem.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Vera Ogette, Joan E. Obi

Acquisition, analysis, or interpretation of data: Vera Ogette

Drafting of the manuscript: Vera Ogette

Critical review of the manuscript for important intellectual content: Vera Ogette, Joan E. Obi

Supervision: Joan E. Obi

Disclosures

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following:

- **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work.
- **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work.
- **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

Data sharing is not applicable. This work is theoretical; no datasets were generated or analyzed.

References

1. Acharya VV, Brunnermeier MK, Pierret D: Systemic risk measures: From the panic of 1907 to the banking stress of 2023. Annual Review of Financial Economics. 2025, 17:1-26. 10.1146/annurev-financial-112823-015828
2. Acharya VV, Pedersen LH, Philippon T, Richardson M: Measuring systemic risk. The Review of Financial Studies. 2017, 30:2-47. 10.1093/rfs/hhw088
3. Adrian T, Brunnermeier MK: CoVaR. American Economic Review. 2016, 106:1705-1741. 10.1257/aer.20120555
4. Adrian T, Covitz D, Liang N: Financial stability monitoring. Annual Review of Financial Economics. 2015, 7:357-395. 10.1146/annurev-financial-111914-042008
5. Aldasoro I, Gambacorta L, Korinek A, Shreeti V, Stein M: Intelligent financial system: How AI is transforming finance. Journal of Financial Stability. 2025, 81:101472. 10.1016/j.jfs.2025.101472
6. Baethge C, Goldbeck-Wood S, Mertens S: SANRA—a scale for the quality assessment of narrative review articles. Research Integrity and Peer Review. 2019, 4:5. 10.1186/s41073-019-0064-8
7. Benoit S, Colliard JE, Hurlin C, Pérignon C: Where the risks lie: A survey on systemic risk. Review of Finance. 2017, 21:109-152. 10.1093/rof/rfw026

8. Beutel J, List S, von Schweinitz G: Does machine learning help us predict banking crises?. Journal of Financial Stability. 2019, 45:100693. 10.1016/j.jfs.2019.100693

9. Billio M, Getmansky M, Lo AW, Pelizzon L: Econometric measures of connectedness and systemic risk in the finance and insurance sectors. Journal of Financial Economics. 2012, 104:535-559. 10.1016/j.jfineco.2011.12.010

10. Bisias D, Flood M, Lo AW, Valavanis S: A survey of systemic risk analytics. Annual Review of Financial Economics. 2012, 4:255-296. 10.1146/annurev-financial-110311-101754

11. Bluwstein K, Buckmann M, Joseph A, Kapadia S, Şimşek Ö: Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. Journal of International Economics. 2023, 145:103773. 10.1016/j.jinteco.2023.103773

12. Board of Governors of the Federal Reserve System: Supervisory Guidance on Model Risk Management (SR 11-7). Federal Reserve SR Letter 11-7/OCC Bulletin 2011-12. 2011, Accessed: June 11, 2026: https://www.federalreserve.gov/boarddocs/srletters/2011/sr1107.pdf.

13. Board of Governors of the Federal Reserve System: Comprehensive Capital and Analysis Review and Dodd-Frank Act Stress Tests: Questions and Answers. 2019, Accessed: June 11, 2026: https://www.federalreserve.gov/publications/comprehensive-capital-analysis-and-review-questions-and-anwers.htm.

14. Board of Governors of the Federal Reserve System: Financial Stability Report — November 2025: Purpose and Framework. 2025, Accessed: June 11, 2026: https://www.federalreserve.gov/publications/november-2025-financial-stability-report-purpose-and-framework.htm.

15. Board of Governors of the Federal Reserve System: SR 26-2: Revised Guidance on Model Risk Management. Federal Reserve SR Letter 26-2 (interagency; also OCC Bulletin 2026-13). 2026, Accessed: June 12, 2026: https://www.federalreserve.gov/supervisionreg/srletters/SR2602.htm.

16. Brownlees C, Engle RF: SRISK: A conditional capital shortfall measure of systemic risk. The Review of Financial Studies. 2017, 30:48-79. 10.1093/rfs/hhw060

17. Buckmann M, Haldane A, Hüser AC: Comparing minds and machines: implications for financial stability. Oxford Review of Economic Policy. 2021, 37:479-508. 10.1093/oxrep/grab017

18. Daníelsson J, Macrae R, Uthemann A: Artificial intelligence and systemic risk. Journal of Banking & Finance. 2022, 140:106290. 10.1016/j.jbankfin.2021.106290

19. Daníelsson J, Uthemann A: Artificial intelligence and financial crises. Journal of Financial Stability. 2025, 80:101453. 10.1016/j.jfs.2025.101453

20. Diebold FX, Yılmaz K: On the network topology of variance decompositions: Measuring the connectedness of financial firms. Journal of Econometrics. 2014, 182:119-134. 10.1016/j.jeconom.2014.04.012

21. Financial Stability Board: The Financial Stability Implications of Artificial Intelligence. 2024, Accessed: June 11, 2026: https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/.

22. Financial Stability Oversight Council: 2025 Annual Report. 2025, Accessed: June 11, 2026: https://home.treasury.gov/system/files/261/FSOC2025AnnualReport.pdf.

23. Financial Stability Oversight Council: About FSOC. 2026, Accessed: June 11, 2026: https://home.treasury.gov/policy-issues/financial-markets-financial-institutions-and-fiscal-service/fsoc/about-fsoc.

24. Flood MD, Mendelowitz AI, Nichols W: Monitoring financial stability in a complex world. Financial Analysis and Risk Management: Data Governance, Analytics and Life Cycle Management. V. Lemieux (ed): Springer, Berlin, Heidelberg; 2013. 15-46. 10.1007/978-3-642-32232-7_2

25. Giglio S, Kelly B, Pruitt S: Systemic risk and the macroeconomy: An empirical evaluation. Journal of Financial Economics. 2016, 119:457-471. 10.1016/j.jfineco.2016.01.010

26. Hanley KW, Hoberg G: Dynamic interpretation of emerging risks in the financial sector. The Review of Financial Studies. 2019, 32:4543-4603. 10.1093/rfs/hhz023

27. Holopainen M, Sarlin P: Toward robust early-warning models: a horse race, ensembles and model uncertainty. Quantitative Finance. 2017, 17:1933-1963. 10.1080/14697688.2017.1357972

28. International Monetary Fund: Chapter 3: Advances in Artificial Intelligence: Implications for Capital Market Activities. Global Financial Stability Report, October 2024. International Monetary Fund, Washington, DC; 2024. 77-105.

10.5089/9798400277573.082.CH003

29. Jakubik J, Vössing M, Kühl N, Walk J, Satzger G: Data-centric artificial intelligence. Business & Information Systems Engineering. 2024, 66:507-515. 10.1007/s12599-024-00857-8
30. Kou G, Chao X, Peng Y, Alsaadi FE, Herrera-Viedma E: Machine learning methods for systemic risk analysis in financial sectors. Technological and Economic Development of Economy. 2019, 25:716-742. 10.3846/tede.2019.8740
31. Monin PJ: The OFR financial stress index. Risks. 2019, 7:25. 10.3390/risks7010025
32. Nyman R, Kapadia S, Tuckett D: News and narratives in financial systems: Exploiting big data for systemic risk assessment. Journal of Economic Dynamics and Control. 2021, 127:104119. 10.1016/j.jedc.2021.104119
33. Office of Financial Research: Legal Entity Identifier (LEI). 2021, Accessed: June 11, 2026: https://www.financialresearch.gov/data/legal-entity-identifier/.
34. Office of Financial Research: OFR Unveils New Hedge Fund Monitor for Public Use. 2024, Accessed: June 11, 2026: https://www.financialresearch.gov/press-releases/2024/07/31/ofr-unveils-new-hedge-fund-monitor-for-public-use/.
35. Office of Financial Research: Short-term Funding Monitor. 2025, Accessed: June 11, 2026: https://www.financialresearch.gov/short-term-funding-monitor.
36. Office of Financial Research: About Us. 2026a, Accessed: June 11, 2026: https://www.financialresearch.gov/about/.
37. Office of Financial Research: Bank Systemic Risk Monitor. 2026b, Accessed: June 11, 2026: https://www.financialresearch.gov/bank-systemic-risk-monitor/.
38. Saguato P: Rethinking the financial stability oversight council. Virginia Law & Business Review. 2022, 16:505-558.
39. Tabassi E: Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1 (National Institute of Standards and Technology). 2023, Accessed: June 11, 2026: https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10.
40. Tang P, Tang T, Lu C: Predicting systemic financial risk with interpretable machine learning. The North American Journal of Economics and Finance. 2024, 71:102088. 10.1016/j.najef.2024.102088
41. U.S. Congress: Dodd-Frank Wall Street Reform and Consumer Protection Act. Public Law 111-203, 124 Stat. 1376. 2010, Accessed: June 11, 2026: https://www.govinfo.gov/content/pkg/PLAW-111publ203/pdf/PLAW-111publ203.pdf.
42. U.S. Department of the Treasury: Managing Artificial Intelligence-Specific Cybersecurity Risks in the Financial Services Sector. 2024a, Accessed: June 11, 2026: https://home.treasury.gov/news/press-releases/jy2212.
43. U.S. Department of the Treasury: Artificial Intelligence in Financial Services: Report on the Uses, Opportunities, and Risks of Artificial Intelligence in the Financial Services Sector. 2024b, Accessed: June 11, 2026: https://home.treasury.gov/system/files/136/Artificial-Intelligence-in-Financial-Services.pdf.
44. U.S. Department of the Treasury: Treasury Releases Two New Resources to Guide AI Use in the Financial Sector. 2026, Accessed: June 11, 2026: https://home.treasury.gov/news/press-releases/sb0401.
45. U.S. Government Accountability Office: Financial Stability Oversight Council: Assessing Effectiveness Could Enhance Response to Systemic Risks. GAO-23-105708. 2023, Accessed: June 11, 2026: https://www.gao.gov/products/gao-23-105708.
46. U.S. Securities and Exchange Commission: Form PF; Event Reporting for Large Hedge Fund Advisers and Private Equity Fund Advisers; Requirements for Large Private Equity Fund Adviser Reporting. Federal Register/SEC Advisers Act Release No. 6297. 2023, Accessed: June 11, 2026: https://www.federalregister.gov/documents/2023/06/12/2023-09775/form-pf-event-reporting-for-large-hedge-fund-advisers....
47. Wang GJ, Chen Y, Zhu Y, Xie C: Systemic risk prediction using machine learning: Does network connectedness help prediction?. International Review of Financial Analysis. 2024, 93:103147. 10.1016/j.irfa.2024.103147

How to cite this article:
Ogette V, Obi J E (September 23, 2026) Artificial Intelligence for Systemic-Risk Monitoring in the U.S. Financial System: A Narrative Review of Methods, Data Sources, and Supervisory Applications. Cureus J Bus Econ 3 : es44404-026-00293-9.
DOI https://doi.org/10.7759/s44404-026-00293-9