• WORKSHOP AT ICAIF '26

IAFM'26: Call for Papers

Interpretability & Alignment of Financial Models

7th ACM International Conference on AI in Finance (ICAIF '26) • Milan, Italy • November 14–15, 2026

SUBMISSION DEADLINE	NOTIFICATION	WORKSHOP DATES	SUBMISSION LINK
October 1, 2026	October 15, 2026	Nov 14–15, 2026	CMT submission portal

SCOPE

Description

Machine learning models are now embedded in the core of financial decision-making: credit scoring, fraud and anti-money-laundering (AML) detection, algorithmic trading, portfolio construction, market-risk estimation, and – increasingly – large language models (LLMs) and agentic systems for research, compliance, and client interaction. In this domain, the *why* behind a prediction is a regulatory obligation (EU AI Act), a risk-management necessity, and a precondition for human trust and adoption.

Until recently, financial ML meant discriminative models paired with XAI: a classifier produced a score, and explainability meant linking that output back to input features via feature attribution, counterfactuals, and concept-based explanations. Today the object of study is different: generative-AI-powered *whole pipelines*, whose safety concerns reach far beyond attribution – hallucination, toxicity, robustness, misuse – and whose transformer-based nature moves the interpretability frontier inward, to **mechanistic interpretability**: sparse autoencoders, circuit-level analysis, and feature-steering.

IAFM'26 bridges two communities that rarely meet: the mature *financial XAI* community and the fast-moving *mechanistic-interpretability* and *AI-safety* communities. Finance is an ideal, high-stakes testbed for both.

Only submissions concerning AI in financial services are in scope.

CALL FOR PAPERS

Topics of Interest

- Post-hoc attribution (SHAP, counterfactuals) in credit, fraud/AML, trading, risk – methods and failure modes
- Concept-based and prototype-based explanations grounded in financial semantics
- Mechanistic interpretability of financial LLMs, foundation models, and agents (SAEs, circuits, steering)
- From XAI to genAI safety: hallucination, toxicity, robustness, and misuse of financial LLMs/agents
- Faithful, robust, true-to-the-model evaluation of explanations; benchmarks and metrics
- Interpretability for regulatory compliance (EU AI Act) and model-risk management
- Human factors: explanations for risk officers, auditors, and customers; user studies
- Fairness–interpretability interplay in financial decisions
- Case studies and lessons learned from production deployments

HOW TO SUBMIT

Submission Guidelines

Submitted papers must follow the submission and formatting guidelines of the host ICAIF conference, according to ACM guidelines. Papers must be written in English and submitted in PDF format.

> **Suggested length:** at least four pages, plus up to two additional pages for references.

Submissions will receive **2–3 double-blind reviews** from the Program Committee. The strongest accepted papers will be given oral slots; remaining accepted papers may be presented as posters. Workshops are not archival: accepted papers will not be part of the ACM ICAIF proceedings, and authors are free to publish their work elsewhere.

ORGANIZERS

Organizing Committee

Alan Perotti LEAD ORGANIZER – Team Leader, Model Explainability and Alignment, Intesa Sanpaolo AI Research.
alan.perotti@intesasanpaolo.com

André Panisson CO-ORGANIZER – Team Leader, Trust, Safety & Resilience, Intesa Sanpaolo AI Research.

Francesco Bonchi SENIOR ADVISOR – Head of Intesa Sanpaolo AI Research.