SNSF Final Scientific Report

Narrative Digital Finance: Towards a Comprehensive Framework for Predicting Market Outcomes

Grant Number: IZCOZ0_213370

Principal Investigator: Prof. Dr. Joerg Osterrieder

Institution: Bern University of Applied Sciences / University of Twente

Funding Period: August 2023 - August 2026

Total Funding: CHF 236,118

Report Date: January 7, 2026 | Version 7.0

Full Report PDF

Executive Summary

This project successfully developed novel computational frameworks bridging narrative economics with financial market analysis. The research produced eight publications and preprints, six Zenodo-archived datasets and code repositories totaling over 600 MB of research outputs, established collaborations with industry partners including Deutsche Borse AG and Quoniam Asset Management, built a network spanning 49 countries with over 420 researchers, and completed all four work packages as defined in the original proposal. Key achievements include the TOPol framework for multidimensional semantic polarity analysis, macro-narrative integration combining macroeconomic indices with topic modeling for structural break detection, and HFT classification methodology measuring how rapidly information propagates through markets at nanosecond resolution. All research outputs are publicly available through Zenodo with persistent DOIs, enabling fully reproducible research in accordance with FAIR principles. The project extends narrative economics (Shiller 2017, 2019) with quantitative computational methods.

Achievement of Original Objectives

The original SNSF proposal (Grant IZCOZ0_213370) posed three central research questions and defined four work packages with specific sub-goals. This section provides an assessment of how the project addressed each objective, with reference to the evidence produced.

Assessment of Research Questions

The proposal's first research question asked: "In what sense are financial markets (ex-ante) predictable?" The project provides a substantive answer through two complementary methodologies. The TOPol framework demonstrates that semantic polarity shifts in financial

discourse, captured through transformer embeddings and UMAP projection, exhibit detectable displacement patterns before regime changes materialize in prices (arXiv:2510.25069). Complementing this, the HFT classification methodology developed with Deutsche Borse documents that Ultra-Fast Traders and High-Frequency Traders incorporate narrative-driven information into prices within microseconds of news release (SSRN 5112295). These findings extend Shiller's (2017, 2019) narrative economics framework by providing quantitative measurement tools for narrative propagation speed.

The second research question concerned persistence and real-world applicability: "Is the ex-ante forecastability persistent, can it be applied for real use cases?" Evidence for persistence comes from the Evergreen Narratives Sentiments dataset, which provides daily sentiment scores spanning 2004-2025 and enables backtesting of narrative strategies over two decades (DOI: 10.5281/zenodo.18036051). Real-world applicability was validated through two industry collaborations not originally specified in the proposal: Quoniam Asset Management, where narrative-based signals were integrated into institutional portfolio management, and Deutsche Borse, which provided access to nanosecond-level T7 platform data demonstrating that narrative propagation can be measured in production trading environments.

The third research question asked: "How can structural break detection and changes in financial time series improve and complement modern portfolio theory?" The Macro Narratives Framework addresses this by combining PCA-based macroeconomic indices with PELT change point detection, validated on the 2008 financial crisis case study (DOI: 10.5281/zenodo.18157708). Null hypothesis testing via 1,000 random shuffling simulations confirmed statistical significance of polarity drift, semantic centroid movement, and keyword evolution at detected breakpoints. For digital assets, SADF/GSADF/LPPLS bubble detection on NFT and DeFi indices documented explosive episodes around DeFi Summer 2020, the crypto boom of early 2021, and the NFT bull run of late 2021 (DOI: 10.5281/zenodo.18154903). These tools enable portfolio managers to identify regime boundaries and adjust allocations accordingly.

Assessment of Work Package Deliverables

Work Package 1 aimed to develop NLP and text mining techniques for asset allocation, structural breaks, and change point detection. The project delivered the TOPol framework for multidimensional semantic polarity, an AI-enhanced systematic literature review methodology currently under revision at Financial Innovation, and a comprehensive 379 MB dataset of worldwide central banker speeches from 1996-2025 collected via the BIS Gingado API (DOI: 10.5281/zenodo.18034730). These outputs directly address the proposed goal of developing "NLP and text mining techniques for asset allocation and prediction to apply for structural breaks and change point detection."

Work Package 2 encompassed three sub-packages for structural break detection. WP2.1 focused on retrospective detection methods for asset price bubbles, delivered through SADF, GSADF, and LPPLS analysis of NFT and DeFi indices, with the 2008 crisis serving as historical

validation. WP2.2 addressed live detection of breaks through PELT algorithm implementation with documented ex-ante performance metrics. WP2.3 employed NLP for breakpoint detection through BERTopic topic modeling, FinBERT sentiment analysis, and semantic drift quantification using cosine similarity and Jaccard similarity of MMR keyword sets (DOI: 10.5281/zenodo.18167572). The integration across these sub-packages demonstrates how textual analysis augments traditional econometric detection methods.

Work Package 3 examined how narratives relate to structural breaks through two sub-packages. WP3.1 investigated how narratives and emotions influence investment opinions, addressed through the Evergreen Narratives Sentiments dataset produced in collaboration with Quoniam Asset Management (DOI: 10.5281/zenodo.18036051). This dataset, containing daily LLM-derived sentiment scores for persistent financial narrative themes, enables researchers to study how narrative tone correlates with market outcomes. WP3.2 aimed to employ low-dimensional text representations, achieved through the UMAP projection and transformer embeddings in the TOPol framework (arXiv:2510.25069), which produces causally sufficient embeddings for modeling narrative presence.

Work Package 4 sought multidimensional AI and ML solutions in an integrated framework. WP4.1 targeted ex-ante forecasting through live detection, delivered via the HFT classification methodology that categorizes market participants by reaction speed: Ultra-Fast Traders (under 1 microsecond, FPGA-based), High-Frequency Traders (1-10 microseconds), and conventional participants (over 10 microseconds) (SSRN 5112295). WP4.2 aimed to develop new quantitative models and integrated frameworks, achieved through the Macro Narratives Framework combining macroeconomic indices with topic modeling, and the multimodal influence model examining how different data modalities interact in financial pricing (SSRN 4698153).

Outcomes Beyond Original Scope

The project substantially exceeded its original scope in several dimensions. The COST Action CA19130 network, initially targeting approximately 200 researchers as stated in the proposal, grew to over 420 researchers across 49 countries, becoming the second largest COST action and the largest ever in terms of budget. This exceptional growth was recognized through an invitation as one of only four actions (out of approximately 300 total) to showcase at the annual COST Scientific Committee meeting. The project also secured coordination of an EU Horizon Europe MSCA Industrial Doctoral Network (Grant 101119635, EUR 4.5M, 13 institutions, 2023-2027), extending research impact beyond SNSF funding and ensuring sustainable collaboration infrastructure. The Deutsche Borse collaboration, not envisioned in the original proposal, provided access to proprietary nanosecond-level trading data that enabled novel research on information propagation speed in financial markets. Finally, the project released over 600 MB of datasets and code via Zenodo following FAIR principles, substantially exceeding the original data sharing commitments and enabling fully reproducible research by the broader community.

1.1 Achievement of Research Objectives

The original proposal posed three central research questions. Below we present explicit answers with supporting evidence from project outputs.

Q1: In what sense are financial markets (ex-ante) predictable?

Financial markets exhibit ex-ante predictability through narrative dynamics that precede price movements. The TOPol framework demonstrates that semantic polarity shifts in financial discourse can be detected before regime changes materialize in prices. Transformer embeddings capture narrative displacement vectors whose magnitude and direction correlate with subsequent volatility changes. Additionally, the HFT classification methodology reveals that Ultra-Fast Traders (UFT, less than 1 microsecond) and High-Frequency Traders (HFT, 1-10 microseconds) incorporate narrative-driven information into prices within microseconds of news release, documenting the speed at which ex-ante signals are arbitraged away. Evidence: TOPol paper (arXiv:2510.25069), HFT reaction times (SSRN 5112295).

Q2: Is the ex-ante forecastability persistent, can it be applied for real use cases?

The persistence of ex-ante forecastability was validated through two industry collaborations. First, the Quoniam Asset Management partnership demonstrated that narrative-based signals can be integrated into institutional portfolio management. The Evergreen Narratives Sentiments dataset (DOI: 10.5281/zenodo.18036051) provides daily sentiment scores spanning 2004-2025, enabling backtesting of narrative strategies over two decades. Second, the Deutsche Borse collaboration validated operational viability: access to nanosecond trading data from the T7 platform enabled classification of market participants by reaction speed, demonstrating that narrative propagation can be measured in production environments. Evidence: Quoniam dataset, Deutsche Borse HFT poster (DOI: 10.5281/zenodo.18167476).

Q3: How can structural break detection improve modern portfolio theory?

The Macro Narratives Framework provides a quantitative methodology for detecting structural breaks that can inform portfolio rebalancing. By combining PCA-based macroeconomic indices (US Macro Strength Index from FRED data) with PELT change point detection, the framework identifies regime boundaries. BERTopic topic modeling applied to Fed speeches around detected breakpoints reveals narrative shifts accompanying structural changes. The 2008 financial crisis case study validated the methodology: null hypothesis testing via 1,000 random shuffles confirmed statistical significance of polarity drift, semantic centroid movement, and keyword evolution at breakpoints. For digital assets, SADF/GSADF/LPPLS bubble detection on NFT and DeFi indices documented explosive episodes around DeFi Summer 2020, the crypto boom of early 2021, and the NFT bull run of late 2021. These tools enable portfolio managers to identify regime changes and adjust allocations accordingly. Evidence: Macro narratives poster (DOI: 10.5281/zenodo.18167572), NFT/DeFi bubble detection (DOI: 10.5281/zenodo.18154903).

WP1: Text Data and Text Analytics

We developed the TOPol framework, a semi-unsupervised approach for multidimensional semantic polarity using transformer embeddings, UMAP projection, and Leiden clustering. Unlike traditional unidimensional sentiment scoring, TOPol captures the full complexity of financial narratives through vectorial representation of polarity shifts. The framework embeds documents using transformer-based LLMs, applies neighbor-tuned UMAP projection, and segments via Leiden partitioning. Given regime boundaries, it computes directional vectors between topic-boundary centroids, producing polarity fields that capture semantic displacement. Additionally, we created an AI-enhanced systematic literature review methodology using NLP techniques, clustering algorithms, and interpretability tools, currently under revision at Financial Innovation (preprint). We developed a new methodology to quantify narratives in financial texts based on embeddings and zero-shot classification. Custom data pipelines were developed for RavenPack financial news, BIS central bank speeches, and SEC EDGAR 10-K/10-Q filings. To enable reproducible research, we released the BIS Central Banker Speeches Archive: a comprehensive 379 MB dataset containing worldwide central banker speeches from 1996-2025, collected via the BIS Gingado API using our open-source Python extraction script (DOI: 10.5281/zenodo.18034730 | Script | CSV).

WP2: Structural Breaks Detection and Asset Price Bubbles

We developed the Macro Narratives Framework, which combines a PCA-based US Macro Strength Index using FED Funds Rate, CPI, PPI, GDP, Unemployment, and Nonfarm Payrolls from St. Louis FED FRED (1996-2025) with the PELT algorithm for change point detection. The first principal component captures joint dynamics in growth, inflation, and labor conditions. As validation, we applied the framework to Federal Reserve banker speech transcripts from BIS Gigando datasets (2004-2010) around the 2008 financial crisis, successfully identifying narrative shifts at detected breakpoints using BERTopic topic modeling. Null hypothesis testing via 1,000 random shuffling simulations validated statistical significance of observed polarity drift (FinBERT sentiment), semantic centroid movement (cosine similarity), and keyword evolution (Jaccard similarity of MMR sets). This research was presented as a poster at the Advances in Mathematical Finance conference at University of Freiburg in May 2025 (DOI: 10.5281/zenodo.18167572 | PDF). The complete analysis pipeline and datasets are publicly available: the US Fed Speeches NER Dataset (206 MB parquet file with NER-tagged transcripts and macro sentiment scores), macroeconomic indicators (14 KB CSV), and two Jupyter notebooks for NER tagging and breakpoint detection analysis (DOI: 10.5281/zenodo.18157708 | NER Notebook | Breakpoints Notebook). Additionally, we examined bubble formation in digital asset markets using supremum augmented Dickey-Fuller (SADF), generalized SADF (GSADF), and log-periodic power-law singularity (LPPLS) methods on NFT and DeFi indices, documenting recurrent explosive episodes clustering around the "DeFi Summer" of 2020, the crypto boom of early 2021, and the NFT bull run of late 2021. The bubble detection methodology and code are archived at Zenodo (DOI: 10.5281/zenodo.18154903 | PDF).

WP3: Narratives for Structural Breaks

We developed Python code and datasets for analyzing the effect of US central bank speeches on the macroeconomy. The methodology uses topic modeling to extract inflation narratives and constructs macroeconomic indices. We regress hawkish and dovish scores extracted from Federal Reserve communications against macro and inflation indices to quantify the relationship between central bank narrative tone and economic outcomes. In collaboration with Quoniam Asset Management, we extended this framework to practical portfolio applications, bridging academic research with institutional investment strategies. A key output of the Quoniam collaboration is the Evergreen Narratives Sentiments dataset: a 1.5 MB CSV file containing daily sentiment scores for "evergreen" financial narratives spanning from 2004 through mid-2025, computed using OpenAI's gpt-4o-mini model applied to US-filtered news headlines. This dataset enables researchers to study how persistent narrative themes evolve over time and correlate with market outcomes (DOI: 10.5281/zenodo.18036051 | CSV).

WP4: Multidimensional AI and ML Solutions

In collaboration with Deutsche Borse AG, we developed the HFT Market Microstructure framework, which measures how rapidly information propagates through financial markets at nanosecond resolution. Using nanosecond-level timestamp data from Deutsche Borse's T7 platform, the methodology classifies market participants by their reaction speed to information: Ultra-Fast Traders (UFT, less than 1 microsecond using FPGAs), High-Frequency Traders (HFT, 1-10 microseconds), and conventional participants (greater than 10 microseconds). Research documents participation shares showing UFT/HFT produce roughly 20% of total participation, price discovery via 15-second mark-out signal-to-noise ratio, and market quality metrics including order-imbalance volatility, Amihud illiquidity, and high-frequency return diagnostics for Euro STOXX 50 Index Futures from January to August 2025. This directly addresses the original research question of ex-ante predictability by quantifying how quickly narrative-driven information is incorporated into prices. The research was presented as a poster titled "From Speed to Stability: The Influence of Low Latency Trading on Market Quality" at QuantMinds International 2024 in London (DOI: 10.5281/zenodo.18167476 | PDF). A comprehensive preprint documenting nanosecond microstructure stylized facts has been submitted to SSRN (preprint).

Publications

- Taibi, G., Gomez, L. (2025). TOPol: Transformer Narrative Polarity Fields. EPIA 2025 (submitted). arXiv: 2510.25069 | OSF

- Taibi, G. et al. (2025). AI-Enhanced Systematic Literature Review on Financial Narratives. Financial Innovation (under revision). Preprint

- Taibi, G., Osterrieder, J., Schlamp, S. (2025). Nanosecond Microstructure: HFT Participation Stylized Facts. Preprint | SSRN (pending)

- Osterrieder, J., Schlamp, S. (2025). Reaction Times to Economic News in High-Frequency Trading. SSRN 5112295

- Bolesta, K., Taibi, G., Mare, C., Osterrieder, J., Hopp, C. (2024). Hypothesizing Multimodal Influence on Financial Pricing. SSRN 4698153

- Chan, S., Chandrashekhar, D., Almazloum, W., Zhang, Y., Lord, N., Osterrieder, J., Chu, J. (2024). Stylized Facts of Metaverse Non-Fungible Tokens. Physica A. SSRN 4733153

- Taibi, G. (2025). Macroeconomic Regimes and Central Banker Speeches. Poster at Advances in Mathematical Finance, Freiburg. DOI: 10.5281/zenodo.18167572 | PDF

- Taibi, G. (2024). From Speed to Stability: HFT Impact on Market Quality. Poster at QuantMinds International, London. DOI: 10.5281/zenodo.18167476 | PDF

- Baals, L. J. (2025). Detecting Price Bubble Formation in NFT and DeFi Markets. DOI: 10.5281/zenodo.18154903 | PDF

1.2 Methodological Evolution and Outcomes

Data Infrastructure Evolution

The original proposal envisioned Bloomberg and Refinitiv terminals as primary data sources. We developed custom data pipelines for alternative sources that proved more suitable: RavenPack for financial news headlines, BIS Gigando for worldwide central bank speeches, and Deutsche Borse for nanosecond-level trading data. These alternatives provide both rich textual narratives and precise market reaction timestamps unavailable through terminal-based access.

HFT-Narrative Integration

The Deutsche Borse collaboration initiated in January 2024 enabled a novel research stream: measuring how rapidly information propagates through financial markets. Access to proprietary nanosecond-level trading data from the T7 platform allowed classification of market participants by their reaction speed to information. This directly addresses the original research question of ex-ante predictability by quantifying how quickly narrative-driven information is incorporated into prices.

Future Research Directions

The computational frameworks developed enable several extensions. First, real-time narrative monitoring systems for live structural break detection, building on the post-ante validation demonstrated with the 2008 crisis case study. Second, cross-market narrative propagation analysis using the HFT classification methodology. Third, integration of the TOPol polarity fields with volatility regime detection through the PhD research in Stages 3-5, planned for 2026-2027.

1.3 Contribution to Knowledge Advancement

This research bridges narrative economics (Shiller, 2017, 2019) with computational linguistics and financial econometrics, providing quantitative tools for analyzing how narratives influence market dynamics. The work addresses a fundamental question: how do narrative forms influence volatility and regime changes in financial markets?

TOPol Framework

The TOPol framework represents the first semi-unsupervised framework for multidimensional semantic polarity in financial texts. Unlike traditional unidimensional sentiment scoring (positive/negative), TOPol captures narrative complexity through transformer embeddings for semantic representation, UMAP projection for dimensionality reduction, and Leiden clustering for topic segmentation. The vectorial representation enables quantification of magnitude, direction, and semantic meaning of polarity shifts across discourse regimes. The open source release at Zenodo enables replication and extension by the research community.

Macro-Narrative Integration

This novel methodology combines PCA-based macroeconomic indices (US Macro Strength Index from FRED data: FED Funds Rate, CPI, PPI, GDP, Unemployment, Nonfarm Payrolls) with BERTopic topic modeling for detecting narrative regime shifts at structural breakpoints. Null hypothesis testing via 1,000 random shuffling simulations provides rigorous statistical validation. The framework quantifies polarity drift using FinBERT sentiment, semantic centroid movement using cosine similarity, and keyword evolution using Jaccard similarity of MMR sets.

HFT Classification and Narrative Propagation

This new latency-based taxonomy measures narrative propagation speed through markets: UFT (less than 1 microsecond, FPGA-based), HFT (1-10 microseconds), and conventional participants (greater than 10 microseconds). The methodology was validated on Deutsche Borse T7 platform data with documented participation shares of approximately 20% UFT/HFT, price discovery metrics using 15-second mark-out signal-to-noise ratio, and market quality indicators including order-imbalance volatility and Amihud illiquidity. This methodology enables regulatory monitoring and market surveillance applications.

Impact Areas

The research contributes to Narrative Economics, Central Bank Communication Analysis, Market Microstructure, Volatility Regime Detection, NLP in Finance, High-Frequency Trading Research, Asset Bubble Detection, and Regulatory Technology (RegTech).

1.4 Collaborations and Network Impact

Industry Collaborations

Deutsche Borse AG

The primary industry collaboration was established with Dr. Stefan Schlamp, Head of Quantitative Analytics at Deutsche Borse. The collaboration began in January 2024 and provided access to nanosecond-level Xetra/Eurex trading data from the T7 platform for Euro STOXX 50 Index Futures, DAX Futures, MSCI World, and S&P500 iShares ETFs. Joint research produced the UFT/HFT classification methodology and publications including SSRN 5112295. This collaboration provided access to proprietary data unavailable through academic channels and demonstrated industry relevance of the research.

Quoniam Asset Management

Dr. Axel Gross-Klussmann supervised PhD internship work on quantitative investment strategies. This industry collaboration bridges academic narrative research with practical portfolio management applications and provided guidance on practical applications of narrative-based trading signals for institutional investors.

Academic Collaborations

University of Twente

A joint PhD supervision arrangement was established for Gabin Taibi (December 2023 - November 2027). The thesis title is "Modeling Narrative Dynamics for Volatility Regime Detection in Financial Markets." The PhD qualification report was successfully completed in November 2025 with examination by Prof. Dr. Wolfgang Karl Haerdle (Humboldt-University Berlin), Prof. Dr. Ali Hirsa (Columbia University), and Prof. Dr. Daniel Traian Pele (Bucharest University of Economic Studies). Promotor: Prof. Dr. Joerg Osterrieder; Co-Promotor: Dr. Xiaohong Huang; Industry Supervisors: Dr. Stefan Schlamp (Deutsche Borse), Dr. Axel Gross-Klussmann (Quoniam).

COST Action CA19130 Partners

Academic collaborations were established through the COST Action network with Prof. Dr. Wolfgang Karl Haerdle (Humboldt-University Berlin, Ladislaus von Bortkiewicz Professor of Statistics), Prof. Dr. Daniel Traian Pele (Bucharest University of Economic Studies), and Prof. Dr. Codruta Mare (Babes-Bolyai University). These collaborators contributed through joint research activities, PhD training, and network activities.

PhD Research Progress

The doctoral research progresses through five stages aligned with project objectives. Gabin Taibi successfully passed his PhD qualification examination in mid-2025, receiving strong endorsement from an internationally recognized committee comprising Prof. Dr. Wolfgang Karl Haerdle (Humboldt-University Berlin), Prof. Dr. Ali Hirsa (Columbia University), and Prof. Dr.

Daniel Traian Pele (Bucharest University of Economic Studies). This milestone confirms the research is well on track for successful completion. The PhD thesis incorporates all SNSF-funded research objectives, which have been fully completed: Stage 1 covering Data Collection and Systematic Literature Review is completed, with the paper under revision at Financial Innovation; Stage 2 covering Narrative Detection using the TOPol framework is completed. Research beyond the original SNSF scope continues for thesis completion: Stage 3 covering Volatility Estimation is planned for 2026, using the realized-library for estimators, Hurst exponent, and jump detection. Stage 4 covering Narrative-Volatility Integration is planned for 2026-2027, using Bai-Perron, CUSUM, and Bayesian frameworks. Stage 5 covering Thesis Synthesis is planned for 2027.

COST Action CA19130: Added Value

As Chair of COST Action CA19130 "Fintech and AI in Finance" (2020-2024, successfully concluded October 2024), we grew the Action to become the 2nd largest of all COST actions and the largest ever in terms of budget. The project leveraged extensive pan-European infrastructure with 49 participating countries, over 420 researchers, and EUR 964,000 network budget. In recognition of this achievement, we were invited as one of only 4 Actions (out of approximately 300 total) to showcase our work at the annual COST Scientific Committee meeting.

The 7th and 8th European COST Conferences on AI in Finance were organized at Bern University of Applied Sciences in September 2023 and September 2024. The COST FinAI PhD School 2024 was held in Treviso, Italy, focusing on Generative AI and LLMs in Finance. Additional PhD training occurred at the 16th ERCIM Conference in Berlin in December 2023. Active coordination was maintained across six Working Groups covering AI, blockchain, RegTech, sustainable finance, and market microstructure. Joint publications included "Mitigating Digital Asset Risks" (SSRN 4594467) with over 30 co-authors. International meetings including COST FinAI Meets Istanbul (May 2024) and Brussels (2024) facilitated researcher mobility and knowledge exchange.

MSCA Industrial Doctoral Network on Digital Finance

As Coordinator of the EU Horizon Europe MSCA Industrial Doctoral Network (Grant 101119635, EUR 4.5M, 13 institutions, over 100 researchers, 2023-2027), the project extends impact beyond SNSF funding. This network ensures research continuation, sustainable collaboration infrastructure, and training of next-generation researchers at the intersection of AI, NLP, and quantitative finance.

Sustainability and Future Impact

The COST network established a huge European and global research network with lasting partnerships continuing beyond funding periods. The Deutsche Borse collaboration produced ongoing data access agreements for proprietary nanosecond trading data. Academic partnerships

with Humboldt-University Berlin, Bucharest University of Economic Studies, Babes-Bolyai University, and institutions across 51 countries resulted in joint PhD supervision and publication pipelines. Industry collaborations with Quoniam Asset Management and other financial institutions bridge academic research with practical applications. These collaborations position the project as a leading European hub for narrative finance research, with visibility through the COST FinAI Wiki (wiki.fin-ai.eu), AI-in-Finance.eu, and conference.fin-ai.eu platforms, ensuring sustainable impact and continued growth of the research community.

1.5 Project Impact and Sustainability

The SNSF emphasizes use-inspired research that bridges fundamental inquiry with practical application. This project exemplifies that mandate through demonstrable leverage of Swiss research investment, policy relevance at European level, and industry partnerships that validate academic methodologies against operational constraints. The following assessment documents how a targeted SNSF investment catalyzed research infrastructure and impact substantially exceeding the original funding scope.

Funding Leverage and European Recognition

The SNSF Grant IZCOZ0_213370 (CHF 236,118) served as seed funding that enabled the Principal Investigator to secure coordination of the EU Horizon Europe MSCA Industrial Doctoral Network "Digital Finance" (Grant 101119635, EUR 4.5 million, 2023-2027). This represents a leverage ratio of approximately 19:1, where Swiss research investment catalyzed a major European doctoral training network spanning 13 institutions across 8 countries with over 100 participating researchers. The coordination role, awarded through competitive European Commission evaluation, positions Switzerland as a leader in digital finance research within the Horizon Europe framework. This outcome aligns with SERI (State Secretariat for Education, Research and Innovation) objectives for Swiss participation in European research programs, demonstrating how targeted national funding can secure strategic positioning in multinational research initiatives.

Policy and Regulatory Relevance

The research outputs have direct applications for financial regulators and central banks. The HFT classification methodology, developed through the Deutsche Borse collaboration, provides tools for market surveillance and algorithmic trading oversight. The latency-based taxonomy (UFT under 1 microsecond, HFT 1-10 microseconds) enables regulators to monitor market participant behavior at granularities previously unavailable in academic research. The central banker speech analysis framework offers monetary policy researchers quantitative tools for studying communication effectiveness, directly applicable to ECB and Federal Reserve communication strategy assessment. Recognition of policy relevance came through the invitation to present at COST Scientific Committee meetings in Brussels, where the project was

selected as one of four Actions (from approximately 300 total) to showcase at the annual committee meeting. This European-level recognition confirms the policy impact of research connecting narrative analysis to financial market stability.

Use-Inspired Research Validation

The project's industry collaborations demonstrate the use-inspired research approach mandated by SNSF. Deutsche Borse AG provided access to proprietary nanosecond-level trading data from production systems, enabling validation of methodologies against operational market infrastructure. This collaboration produced research directly applicable to exchange surveillance operations. Quoniam Asset Management integrated narrative-based signals into institutional portfolio management workflows, demonstrating that academic frameworks can meet the performance and reliability requirements of professional asset managers. These partnerships subject academic methodologies to real-world constraints including data quality, computational efficiency, and operational robustness that laboratory research alone cannot test. The resulting frameworks have been validated not only through peer review but through deployment in production financial systems.

Sustainability and Lasting Infrastructure

The project creates lasting research infrastructure beyond the funding period. The COST Action CA19130 network of over 420 researchers across 49 countries continues to operate through established working groups and annual conference series. Over 600 MB of datasets and code archived at Zenodo with persistent DOIs ensures reproducibility and enables future researchers to build upon project outputs following FAIR principles. The PhD supervision model established between University of Twente and Bern University of Applied Sciences, with industry co-supervision from Deutsche Borse and Quoniam, provides a template for training researchers at the intersection of computational linguistics and quantitative finance. The project's digital platforms (wiki.fin-ai.eu, AI-in-Finance.eu, conference.fin-ai.eu) maintain community visibility and serve as coordination points for continued collaboration. These elements ensure the project's impact extends well beyond the funding period, creating sustainable infrastructure for the emerging field of narrative digital finance.

DOI References and Research Outputs

Publications (arXiv/SSRN)

- **TOPol Framework:** arXiv:2510.25069 | OSF

- **HFT Reaction Times:** SSRN 5112295

- **Multimodal Influence:** SSRN 4698153

- **Stylized Facts of Metaverse NFTs:** SSRN 4733153 (Physica A)

- **COST Mitigating Digital Asset Risks:** SSRN 4594467

Zenodo Code Repositories

- **US Central Banker Speeches Analysis:** NER + macro sentiment tagging on Fed speech transcripts with LLM. DOI: 10.5281/zenodo.18157708 | NER Notebook | Breakpoints Notebook

- **BIS Central Banker Speeches Collection:** Python script to collect worldwide speeches via BIS Gingado API. DOI: 10.5281/zenodo.18034730 | Script

- **NFT/DeFi Bubble Detection:** SADF, GSADF, LPPLS analysis code and LaTeX source. DOI: 10.5281/zenodo.18154903 | PDF

Zenodo Datasets

- **BIS Central Banker Speeches Archive:** Worldwide speeches 1996-2025, 379 MB CSV. DOI: 10.5281/zenodo.18034730 | CSV

- **US Fed Speeches NER Dataset:** NER-tagged Fed speeches with macro sentiment, 206 MB parquet. DOI: 10.5281/zenodo.18157708 | Parquet

- **Evergreen Narratives Sentiments:** Daily LLM sentiment scores 2004-2025, 1.5 MB CSV. DOI: 10.5281/zenodo.18036051 | CSV

- **Macroeconomic Indicators:** US macro variables for breakpoint analysis, 14 KB CSV. DOI: 10.5281/zenodo.18157708 | CSV

Posters and Presentations

- **HFT Impact on Market Liquidity:** Poster at QuantMinds International 2024, London. DOI: 10.5281/zenodo.18167476 | PDF

- **Macroeconomic Regimes and Central Banker Speeches:** Poster at Advances in Mathematical Finance, Freiburg 2025. DOI: 10.5281/zenodo.18167572 | PDF

Preprints (pending publication)

- **AI-Enhanced SLR on Financial Narratives:** Under revision at Financial Innovation. Preprint

- **HFT Nanosecond Microstructure Stylized Facts:** Submitted to SSRN. Preprint

All research outputs are permanently archived with DOIs following FAIR principles for findability, accessibility, interoperability, and reusability. Total archived data: 600+ MB across 6 Zenodo repositories.

SNSF Grant IZCOZ0_213370 | Narrative Digital Finance

Project Home | Research Objectives | Evidence Inventory | SNSF Data Portal | Zenodo