Institut Louis Bachelier

DEFINING & MEASURING THE IMPACT OF AI ON RISK MANAGEMENT & THE RELIABILITY OF FINANCIAL MODELS

Workshop by Institut Louis Bachelier, GENCI and AI Factory & Finance Vertical, at Palais Brongniart, Paris, the 3rd of February 2026.

WORKSHOP SUMMARY: THE CHALLENGE OF BENCHMARKS AND EUROPEAN SOVEREIGNTY IN AI FOR FINANCE

The Institut Louis Bachelier (ILB), in partnership with GENCI, held a workshop at the Palais Brongniart dedicated to the "IA Factory" and the challenges of AI in finance. This meeting brought together researchers, institutions, and market participants with a unified goal: identifying the sector's needs for reference datasets, benchmarks, and metrics tailored to financial realities. The discussions were particularly rich, followed by numerous questions from the audience, signaling the ecosystem's strong interest in these topics. Finance is among the sectors most impacted by AI, which reinforces the necessity for robust and shared evaluation standards.

In the opening session, Marie Brière (General Director, ILB) and Cédric Auliac (GENCI) reminded attendees that AI is already present at every level of financial decision-making—including compliance, fraud detection, forecasting, allocation, risk analysis, and customer relations. Cédric Auliac specifically noted that while GENCI's computing power has traditionally served nuclear physics and fluid dynamics, finance has now become a vital "vertical" for these national resources. The priority has shifted toward reliability, measurement, and evaluation frameworks.

The speakers addressed fundamental questions: What defines a reliable AI model in finance? How can we measure its precision and robustness against regime changes or adversarial attacks? How do we measure biases, and how can we evaluate its global added value in a multidimensional approach that includes financial and environmental costs? Furthermore, the workshop raised vital questions regarding sovereignty and independence: must we fine-tune and evaluate these models specifically on European data?

The roundtable featured a deep exchange between researchers and practitioners from BNP Paribas Global Markets, Natixis CIB, ARDIAN, Dragon LLM, École Polytechnique, and Google DeepMind. These discussions highlighted a significant gap: while financial use cases for Large Language

Models (LLMs) already exist, the evaluation frameworks are not yet sufficiently structured. The primary bottleneck is not the application itself, but the measurement of reliability. Similarly for risk and valuation applications of machine learning like deep hedging or fast pricing, experts argued that industrials must impose rigorous scientific standards, requiring "evaluation sets" for every use case before production. Human verification remains central, with a clear requirement for continuous supervision and validation to ensure the AI remains a "productivity helper" rather than an autonomous risk.

The speakers also prioritized frugality and targeted approaches: choosing the right model for the right use and prioritizing methodology, traceability, and control. They warned against in sample tests and qualitative assessment over out of sample comparison on clear benchmarks. For texts, this includes using small, specialized models, which can outperform generalist giants on specific financial tasks while remaining affordable and energy-efficient.
The round table covered topics of operational productivity, of valuation, hedge and risk management, and of models evaluation in presence of distributional shift, concluding on sovereignty aspects.

The shared conclusion is that the goal is no longer just to "make models work," but to prove, measure, and frame their behavior within a financial environment. All participants stressed the importance of developing structured European benchmarks to ensure sovereignty. This dynamic requires a collective effort to pool data, metrics, and methods through cooperation between research and industry. Consequently, a working group is being established in partnership with the Institut Louis Bachelier, the IA Finance vertical of the IA Factory and GENCI, and the Cercle IA et Finance to advance these sovereign standards.

Overview

1. Introduction by Marie Brière (Institut Louis Bachelier)

- AI has become a **major lever in finance**, supported by unprecedented levels of data and computing power, and is now utilized across all decision-making levels—from regulatory compliance and fraud detection to portfolio construction and customer interaction.
- The rapid adoption of AI raises critical questions regarding **reliability and evaluation**, specifically how to measure models' accuracy, detect and mitigate biases, ensure robustness against adversarial attacks
- Widespread use across financial institutions increases verification needs: should we assign humans to these costly tasks? LLMs as judges ? How can effective oversight be organized?
- Academic evaluation shows that general purpose LLMs often perform poorly on financial tasks. This suggests the need for models **finetuned** on financial corpora and for **finance-specific benchmarks**
- **Sovereignty** and independence issues arise: should models be finetuned and assessed on regional (e.g. European) data to avoid external dependencies and ensure policy alignment?
- The immediate challenge is to define relevant **financial use cases, associated benchmarks and metrics** that together allow rigorous evaluation and governance of deployed models.

2. Introduction by Cédric Auliac (GENCI)

- GENCI coordinates the French node of **"IA Factory"**, a European program which aims to provide seamless access to key resources - ranging from high-performance computing infrastructures and tailored dataspaces to technical support or end-users training- to accelerate public and private AI innovation in a broad range of application domains, including **sovereign financial research**.

3. Challenges for AI in risk management

Talk by Charles-Albert Lehalle (CMAP, Ecole Polytechnique)

Charles-Albert reminds that this Workshop was the kick-off of the finance vertical of the IA Factory project. The goal of this project, selected by the European Commission is to focus on two fields:

- **Risk management and stress test** of large (and nonlinear) portfolios: Risk management is a permanent concern in the financial industry, it is of course enforced and improved by natural authorities and European bodies (ESMA, EBA). Risk management drives most decisions in the industry.
- **Trustable GPTs for finance**: LLM which encompass both textual data and tokenized time-series addressing critical challenges such as look-ahead bias and efficiency in handling time series data.

Along his talk, he mentions that:

- A core objective of the “finance vertical” of IA Factory is to **reduce fixed costs for innovation**, allowing startups and researchers to access sovereign AI supercomputers and innovation services required for modern AI development.
- The financial sector suffers from a **lack of standardized benchmarks and datasets**; unlike other fields that progressed through famous shared datasets (like the "iris" flowers), finance lacks these "stylized cases" to stimulate competitive research.
- Non-LLM AI applications in finance are categorized into three families:
    - **portfolio construction**: concentration on an investment thesis and diversification among bets, account for transaction costs and out of sample risk;
    - **hedging and risk replication** for derivatives: using scientific machine learning or direct application of ML or AI to solve approximately associated optimisation problem, adding more risk factors (regulation), and liquidity constraints
    - and **optimal liquidation/market making**: join liquidity measures (bid-ask spread, quantities) to prices, and become discrete (at the level of messages rather than of price returns); adopt policies that can switch between endogenous and exogenous phases of price formation.

4. Generative AI, bias, and truthfulness for market finance

Talk by Damien Challet (CentraleSupelec)

- LLMs face a significant **"forward-looking bias"** because they are trained on data up to a specific date; this makes traditional financial backtesting difficult, as it is hard to distinguish if a model is "thinking" or simply **memorizing future events** like the COVID-19 pandemic.
- Mitigation strategies include **recursive prompting**, such as repeating a question multiple times (up to 50) within the prompt to better select the correct "subspace" of response, or using models specifically fine-tuned on "point-in-time" data.
- Research indicates that **internal "hesitation" in token activation** can be measured; when a model is about to make an error, its internal probability distributions shift, potentially allowing for an internal measure of **model confidence**.

5. Roundtable Panel

Moderated by **Marie Scheid** (Ecole Polytechnique), the panelists addressed key points needed to define finance-specific use cases, benchmarks and use cases.

The Panelists:

- Gaëtan Caillaut (Researcher, Dragon LLM)
- Laurent Carlier (Head of AI LAb, BNP Paribas Global Market)
- Alain Durmus (Researcher,Ecole Polytechnique)
- Romuald Elie (Researcher, Google DeepMind)
- Pascal Oswald (Head of Risk Methodologies, Natixis)
- Guillaume Rigaud (Data Scientist, Ardian)

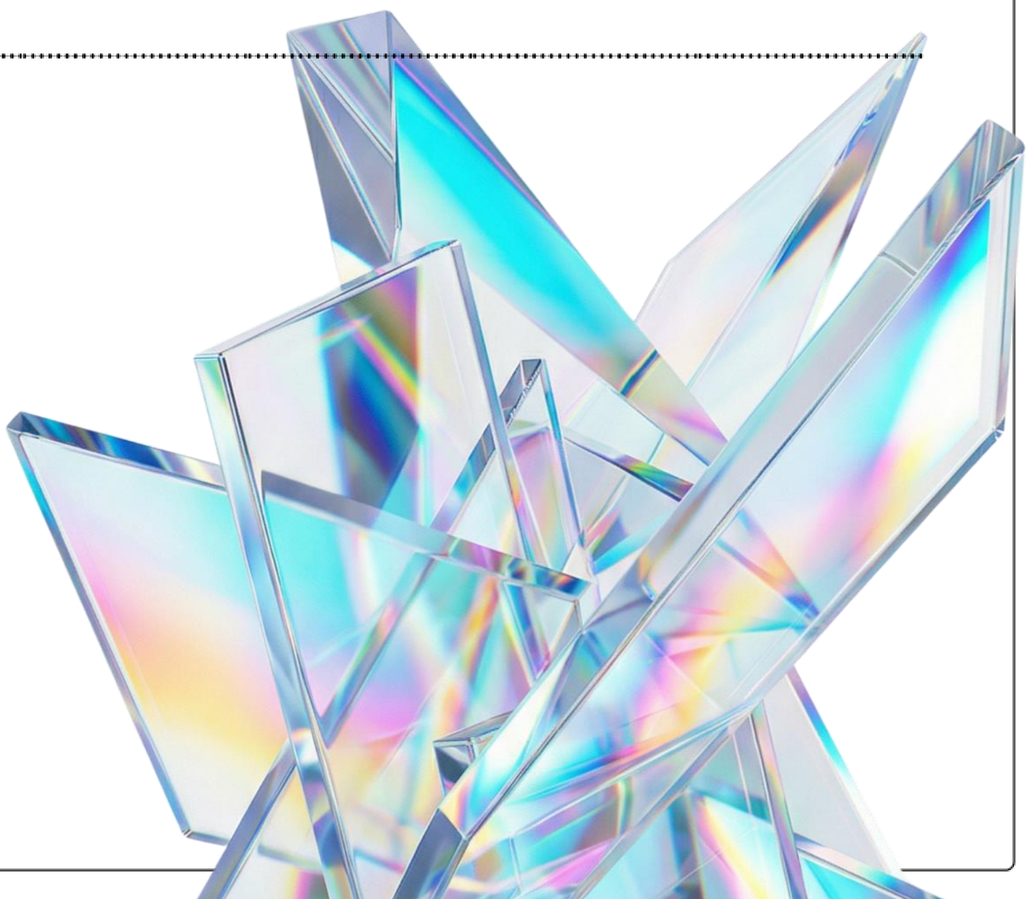
The main three themes addressed were:

1. **Operational Productivity:** Operational productivity in finance utilizes LLMs to automate six functional areas—including data extraction, report generation, and assisted coding—while transitioning toward autonomous agents for complex workflow automation. To ensure security and economic frugality, firms are deploying specialized, small-scale models on private

platforms and performing cost-benefit analyses to replace manual tasks like legal document reviews.

2. **Valuation, Hedging and Risk Management of financial products:** AI optimizes financial risk management and pricing by replacing computationally expensive Monte Carlo simulations and biased approximations with near-instant "substitution models" that enable faster, more accurate evaluations of complex products. To establish industry-wide reliability, experts emphasize the necessity of public reference datasets—leveraging synthetic data from standard models—and federated metrics to create a consensus for characterizing and comparing model performance.
3. **Evaluation of models, From Distributional Shift to Sovereignty:** Effective financial AI evaluation requires a transition from static benchmarks to federated metrics and use-case-specific "evaluation sets" capable of detecting non-stationary distributional shifts and "forward-looking" training biases. To ensure technological and financial sovereignty, the sector must develop independent European benchmarks and federated reports to mitigate the US-centric cultural and linguistic biases inherent in generalist models.

Plus meaningful remarks on **Economic and Technical Frugality:** Industrials prioritize the **least expensive model** that meets accuracy requirements, often preferring specialized, smaller models (like Bert or Mamba architectures) over generalist giants to manage inference costs and latency.



1. OPERATIONAL PRODUCTIVITY

Operational productivity is a primary driver for the adoption of Large Language Models (LLMs) in the financial sector. The sources describe it as a **lever to facilitate the daily work of operators** by automating repetitive or data-heavy tasks.

The Six Key Categories of Productivity Use Cases

Laurent Carlier outlines six broad functional areas commonly observed in the operational use of LLMs:

- Q&A on Documents: Using chatbots to retrieve information from vast, unstructured document databases via natural language.
- Q&A on Structured Data: Laurent described this area as as the "little brother" of document Q&A, this use case involves "interrogating structured data" such as "databases of market data, trade data, or economic figures" to facilitate data access
- Text Transformation: Handling low-value-added manual tasks like rewriting emails, translating documents, and generating meeting minutes or action points.
- Text and Report Generation: Automatically creating market commentaries, final reports, and professional presentations from structured data. Examples include "generating a market commentary from information, generating a final report, or generating a presentation".
- Data Extraction: Structuring previously unstructured data to accelerate the deployment of specific financial use cases.
- Assisted Coding: This is described as a "revolution" in how developers and data scientists write code compared to just a few years ago.

Beyond these six categories, Laurent Carlier identifies a "supplementary layer" currently in development: agents. Unlike the productivity tools described above, agents are focused on action and automation of workflows, allowing the LLM to interact with databases or trigger predefined actions within strictly controlled workflows (for example, sending notifications or preparing transactions subject to human validation).

The use case of “term sheets”

Laurent Carlier underlines that Term Sheets are examples of documents that are comparatively more standardized in structure, alongside fund prospectuses, annual reports, and KYC (Know Your Client) documents.

Laurent notes that many financial institutions perform similar, repetitive processing tasks on such documents (for instance, identifying specific clauses or indicators), which creates an opportunity to reflect on methodological approaches.

He suggests that while market or client data is a proprietary competitive advantage, the processing of highly standardized documents such as term sheets raises similar operational challenges across institutions.

The objective of applying AI to such documents is to support operators by automating repetitive extraction tasks, allowing teams to focus on higher value analysis and oversight.

Specialized Industry Applications

On general productivity, **Guillaume Rigaud** reminded that having a secured and private platform for Gen AI is a requirement to any financial use case, to ensure confidentiality, especially in Private Equity. Beyond general productivity, **Guillaume Rigaud** underlined specific roles in the **investment industry** highlight targeted operational gains, using LLM to speed up processes while keeping human in he loop for the final review:

- **Legal and Compliance:** LLMs are used to review legal documents like NDAs. By detecting potential breaches of template clauses, the firm can reduce the number of costly exchanges required with outside counsel.
- **Private Equity & Investor Relations:** Industrials use LLMs to "crunch" private databases and draft responses to investors.

The Developer and Researcher Perspective

Gaëtan Caillaut (Dragon LLM) Necessity of Fine-Tuning: "*we are going to need at some point... to fine-tune these models... because our generalist models, we will realize they are not good on very sharp cases*"

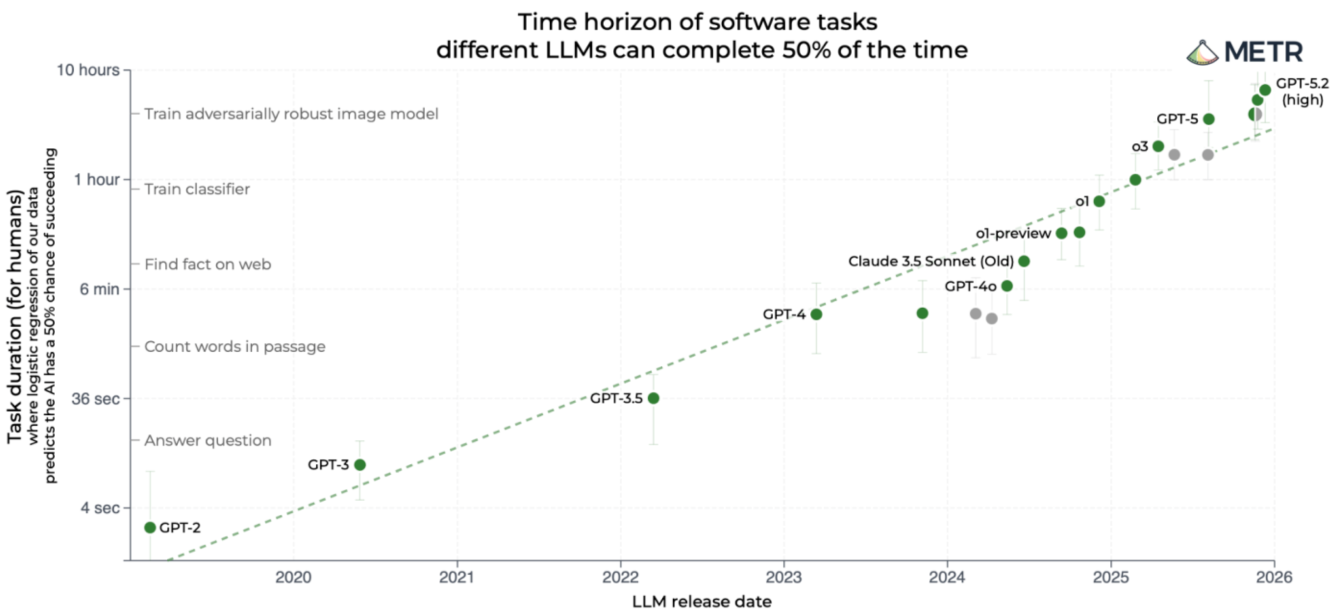
Panelists from **LLM builders** and **academic research** emphasize productivity in the creation and maintenance of these tools:

- **Gaëtan Caillaut** focuses on creating small, specialized models for finance: these are designed to be "affordable" for firms with limited computing power, focusing specifically on information extraction and summarization.
  Benchmarks are crucial during the distillation because they focus the preserved functionalities of the LLM on a clear list of use cases.
- **Romuald Elie** highlights the use of LLMs as a "thought partner" to brainstorm and assist with complex mathematical proofs from a **research** standpoint.

Operational Constraints and associated metrics

The panelists discussed elements that should allow to qualify new tools:

- **Efficiency Ratio:** For a task to be worthwhile, it must be faster for a human to verify the LLM's output than to perform the task from scratch.



The progress of tasks performed by LLM in terms of their duration (y-axis) through time (x-axis).

Source: metr.org the 2025-03-19

- **Frugality: Laurent Carlier** mentions internal frugality principles " whereby industrial players seek to rely on the least complex and least costly model that is fit for purpose (e.g., using an older, smaller BERT model for simple classification) rather than a costly LLM. **Guillaume Rigaud** notes that frugality is also measured by a direct **cost-benefit analysis**; for example, the cost of an LLM inference can be compared against fees saved by reducing (but not replacing) requests to outside counsel, keeping human in the loop for final decision or review of an output.

- **Security:** To prevent data leaks to public AI, firms like Ardian or BNP Paribas (in partnership with Mistral AI) rely on secure internal environments and private platforms to enable the use of LLMs **without exposing sensitive investment data**.

2. VALUATION, HEDGING AND RISK MANAGEMENT OF FINANCIAL PRODUCTS

Romuald Elie (Deepmind) on **The Goal of Comparison** "*Benchmarks are fabricated to be able to compare models with each other... the objective of a benchmark is not at all to be perfect... but that it be diverse and complete the set of existing benchmarks*"

In the financial sector, the application of AI to **risk and pricing efficiency** focuses on replacing computationally expensive traditional models with fast, data-driven alternatives and refining the precision of risk measures. This is referred to as: **Deep hedging**, **fast pricing**, and **scientific machine learning** (from a more general perspective).

Extending numerical solvers for risk management with AI

Pascal Oswald explains that current risk monitoring produces vast amounts of data but requires **massive computing power**. Because they cannot produce every measure perfectly, banks use approximations (like Taylor expansions), which unfortunately introduce **biases** into their risk monitoring.

Charles-Albert Lehalle notes that for products like derivatives, banks must solve complex optimization problems using **Monte Carlo simulations** to find optimal hedging strategies along future price trajectories. These simulations are highly resource-intensive, learning from existing simulations the direct mapping between risk factors and current market conditions to prices, values, and optimal hedge would spare a lot of computing power and allow **more frequent risk management** that could take into account more parameters. **Romuald Elie** suggests that **Generative AI** and **Reinforcement Learning** will likely partially replace or augment traditional Monte Carlo methods for pricing and hedging within the next 5 to 10 years.

Pascal highlights the primary efficiency gain: while a traditional physical model might take **30 seconds** to price a complex product, a trained AI **inference model** can provide a result in **milliseconds**. This allows for much faster and more frequent risk evaluations.

Cédric Auliac supports this claim with an analogy of what GENCI is seeing in various scientific domains: In Computational Fluid Dynamics or particle physics for instance, AI can be used to train "surrogate models" that approximate complex physical / numerical models and generate results —albeit slightly degraded— much faster than computational heavy simulations. However, for complex problems, the computational cost of putting together large and diverse simulated datasets to train large surrogate models requires cooperation.

Laurent Carlier advocates for the creation of **common libraries of synthetic data**—artificial environments based on statistical models (Heston, etc.)—to train large scale reference models for financial time series. **Romuald Elie** agrees, comparing the approach to training reinforcement learning on Atari games; by creating synthetic **market microstructure environments** (like order book simulators), the community can test algorithms without risking real capital.

Part of the solution: having reference datasets and associated metrics

Alain Durmus (Ecole Polytechnique) on **The Metric Deficit in Finance**: "*What surprised me in starting Marie's thesis was the lack of metrics on how to evaluate our models... we can't find a good metric for financial time series to really characterize the benefit of one model over another*"

Pascal Oswald suggests that for such applications, there is no real difficulty to get as much data as it is needed since they come from Monte-Carlo simulations. It means that having

- a set of generative models based on standard models of the industry used in Monte-Carlo or finite difference solvers,
- associated large datasets of Monte-Carlo generated trajectories associated to standard sets of parameters,
- a list of standard derivative products and their valuation for some values of their parameters (strikes, maturities, etc)

could be public and **would allow fair comparisons** between machine learning and AI models. This allows for precise calibration in a controlled environment where the "truth" is known, bypassing the need for restricted and expensive data from providers like Bloomberg.

Cédric Auliac points out that machine learning can also be used for **inverse optimization**, effectively parametrizing an ideal physical model to better fit real-world market data. Adding real data would hence be a plus (and would allow to compute risk metrics), nevertheless **Marie Brière** underlines that, **historical times series of datasets**, that are not public and in general are expensive, would be needed.

Alain Durmus emphasizes that a benchmark is only useful if it is paired with a **metric** that federates the community. He mentions that the financial applications are far behind other domains (like images or texts): currently, there is no consensus in the industry and in academia on how to properly characterize the benefit of one financial model over another. It is very important to build such a consensus. **Charles-Albert Lehalle** notes that while fields like image processing evolved through famous datasets like "Iris" or "MNIST," finance lacks these "stylized cases" to stimulate competitive research.

Gaëtan Caillaut concurs on the lack of adequate benchmarks, even for language models. He points out that existing benchmarks like "FinBert" focus more on general financial jargon than on the **specialized tasks** required by industrials, such as processing Swift messages.

3. EVALUATION OF MODELS: FROM DISTRIBUTIONAL SHIFT TO SOVEREIGNTY

Speakers from both academia and industry argue that datasets alone cannot drive innovation; they must be paired with clear, federated metrics to quantify model superiority. **Alain Durmus** makes clear that a benchmark is essentially useless without a metric that allows the community to characterize exactly why one model is better than another. He notes that unlike image processing, where "blurriness" is visible, financial time series require specialized quantification that remains an open question. He identifies **non-stationarity** and **multimodality** as the primary technical challenges that make finding these financial metrics so difficult.

The "*most important thing is not so much the benchmark as the evaluation set*" specific to each use case, insists **Laurent Carlier**. He stresses that performance must translate into "business performance," such as the Sharpe ratio for investment strategies. **Pascal Oswald** adds that for risk and pricing, the primary metrics are accuracy and computation time—for example, replacing an 8-hour Monte Carlo simulation with an inference model that provides a result in milliseconds.

Distributional Shift and Non-Stationarity

Marie Brière warns that financial data is **non-stationary**; a reference dataset created today could become obsolete within a year if a new market crisis occurs. Indeed financial markets are characterized by constant "distributional shifts," meaning a model's performance can degrade rapidly as market regimes change.

Pascal Oswald suggests that once a model is in production, firms must implement **"sensors"** to detect the exact moment a model begins to **"derail"** due to a change in market regime, whereas **Romuald Elie** suggests a more structural solution: treating the market as a **multi-agent system** from the start of the design phase to better model and evaluate these non-stationary shifts. Simulators could be a way to study these dependencies. He adds that the **non-stationarity** of financial data means a benchmark created today might be obsolete in a year if a market crisis occurs, requiring constant updates.

The Role of Evaluation: In-Sample and Out-of-Sample Challenges

Beyond non-stationarity, a difficulty in evaluating financial AI is the **"forward-looking bias,"** where a model appears to perform well because it has already "seen" the future during its training phase. **Damien Challet** identifies this as the primary challenge for LLM backtesting. If an LLM is trained on data up to a specific date, any "out-of-sample" test before that date may just be "*memorization*" rather than "*reflection*". **Guillaume Rigaud** echoes this, stating that because LLMs are "black boxes," it is nearly impossible to be truly **"point-in-time"**; one can never be certain that the evaluation documents weren't already part of the training corpus, unless applying very rigorous evaluation with documents published after the release of the model.

The banking industry approaches **the difference between benchmark and evaluation** through the lens of operational validation. Moreover, the panel underlines that the non-stationarity of the data may require the benchmarks to be regularly updated, or at the extreme that the regulators own a “hidden” extension of the benchmarks, for stress testing. **Gaëtan Caillaut**, in accordance with **Alain Durmus** and **Romuald Elie**, nevertheless mentions that benchmarks are not meant to be the ultimate goal, but point to stylized use cases with stylized metrics, to show where the model's efficiency should focus.

Laurent Carlier argues that for the industry, a **specific "evaluation set"** for each use case is more critical than a general benchmark. He emphasizes that projects moving toward production typically require a validated evaluation dataset, often on the order of several hundred synthetic or manually tagged questions, to ensure scientific rigor.

Marie Brière’s remark on the need for a **multidimensional evaluation** of AI that includes not just accuracy, but also its environmental footprint and financial overhead, marks the transition of the discussion to a more general concern: sovereignty.

The Challenge of European Sovereignty

The discussion quickly shifted towards the need for **sovereign benchmarks** that reflect European financial realities rather than relying on US-centric models, **Marie Brière** raising the fundamental question of whether models must be **fine-tuned and evaluated specifically on European data** to ensure independence.

- **Cédric Auliac** notes that France is one the few European nations to have selected "Finance" as a specific vertical for its **IA Factory**, highlighting a strategic move toward **technological and financial sovereignty**.
- **Gaëtan Caillaut** highlights a linguistic and cultural bias: most current benchmarks are in English, which fails to capture the specificities of the French or European markets.

Roundtable participants warn that generalist LLMs (like ChatGPT) often exhibit a **US-centric bias**, such as recommending US ETFs over European investments. To counter this, they suggest federating **European financial reports** (like those found in the "EDGAR" equivalent) to ensure that European investment theses and company specificities are respected.

Appendix (on Economic and Technical Frugality)

Economic and technical frugality is a central theme in the sources, as speakers emphasize that the adoption of AI in finance must be balanced against its high financial and environmental costs.

Laurent Carlier notes that while setting up an LLM use case might take only 30 seconds, the long-term cost of production often makes it worth the extra time to properly train a smaller, more specialized model that can **reach very high accuracy levels on well-defined tasks at a fraction of the compute cost**.

1. Specialized and "Small" Language Models

- **Gaëtan Caillaut** focuses on creating **small-scale, specialized LLMs** (e.g., much smaller than 70 billion parameters) because many financial institutions do not have the massive GPU resources required to run giant models.

- The goal is to produce models that are **"abordables" (affordable)** and run efficiently on low-capacity hardware while maintaining high performance on specific financial tasks.
- **Alain Durmus** highlights research into **“model compression”**, which attempts to reduce a model's size (e.g., from 70 million to 1 million parameters) to gain speed in inference and calculation.

2. Architectural Innovations for Efficiency

- **Romuald Elie** discusses **"distillation,"** a process where the qualities of a large, expensive model are replicated into a smaller one that can run cheaply on portable devices like phones.
- To manage the high cost of the "attention mechanism" in traditional Transformers (which requires re-reading all previous text for every new word), **Gaëtan Caillaut** mentions exploring new architectures like **"Mamba"** which are designed to be more frugal.
- **Romuald** also points to **"Mixture of Experts" (MoE)** architectures, which create "shortcuts" so that the model does not have to calculate all its weights for every single query, significantly reducing the computational load.

MARIE BRIÈRE

Managing Director | Institut Louis Bachelier

Marie Brière, PhD, is the Managing Director of Institut Louis Bachelier and Head of the Investor Intelligence and Academic Partnership at Amundi Investment Institute. She conducts research on portfolio choice, with a focus on sustainable and household finance, pensions and the impact of financial technology, advising the strategic decisions of institutional investors and the design of investment solutions for retail clients. Marie is the Chairman of Inquire Europe, Chairman of the Scientific Committee of the European Savings Observatory, a member of the expert group advising the ESMA Risk Standing Committee and a member of several scientific councils, such as the French supervisory authority ACPR, the European Capital Market Institute of CEPS and the Principles for Responsible Investment (PRI). She is also a senior associate researcher with Université Libre de Bruxelles and Paris Dauphine PSL University and serves as Editorial Board member of the Financial Analyst Journal. A frequent speaker at international conferences, her scientific articles have been published in academic journals and her work has been featured in several news outlets including the Financial Times and the Wall Street Journal. Marie holds a PhD in Economics from Université Paris X and graduated from ENSAE.

ilb Institut Louis Bachelier

CÉDRIC AULIAC

AI Program Director | GENCI

Cédric Auliac leads GENCI's artificial intelligence program. After earning his PhD in computational biology and biostatistics in 2008, he joined CEA-LIST (Institute for Intelligent Digital Systems, Saclay) as a senior AI researcher and data scientist. After completing an MBA at HEC Paris in 2016, he took responsibility for industrial partnerships and business development at CEA-LIST, and then, in 2020, he became head of the "Signal Processing & Machine Learning" research unit. In 2023, he joined CEA's Research and Technology Division to oversee the AI program, and also joined the national ASIC program agency to help establish research programs on AI hardware and computing infrastructure. Cédric Auliac currently coordinates the "French IA Factory," a national platform designed to accelerate the adoption of AI.

GENCI HPC at the service of knowledge

CHARLES-ALBERT LEHALLE

Professor | École Polytechnique

Charles-Albert Lehalle is a professor at École Polytechnique in Paris, where he teaches and conducts research on liquidity, price formation, and the use of artificial intelligence in financial markets. Previously, he served as Global Head of Quantitative Research at Crédit Agricole Cheuvreux and Head of Quantitative Research on market microstructure at Crédit Agricole Corporate and Investment Bank, before joining Capital Fund Management (CFM) for seven years. Winner of the Europlace Institute for Finance (EIF) Best Paper Award in Finance in 2016, he has published more than eighty articles and book chapters analyzing the key features of modern markets. He is also a member of the scientific committee of the Institut Louis Bachelier and teaches at UC Berkeley, Paris 6 Sorbonne University, and in the "Probabilities and Finance" Master's program at École Polytechnique.

École Polytechnique | Institut Polytechnique de Paris

DAMIEN CHALLET

Associate Professor - Researcher | CentraleSupélec

Damien Challet is a researcher and lecturer at Université Paris-Saclay, CentraleSupélec, within the Laboratory of Mathematics and Computer Science for Complexity and Systems, and co-head of the FiQuant team. He is also Scientific Advisor to the EconophysiX Lab.

His research interests focus on learning and inference in high-dimensional systems, including market microstructure, predictive relationships, and agent-based models.

CentraleSupélec | université PARIS-SACLAY

LAURENT CARLIER

Head of the AI Lab | BNP Paribas GM

Laurent Carlier, master in Applied Math from Centrale Paris and master in quantitative finance at Sorbonne University.
More than 25 years of experience in banking industry, first as head of various quantitative research teams in London, Tokyo and New York, and since 10 years leading from Paris the Data and AI Lab for Corporate and Investment Banking – Global Market at BNP Paribas.

Main focus is to deliver projects to automate tasks, increase productivity or improve pricing and risk management of financial products thanks to both Traditional and Generative AI.

BNP PARIBAS

MARIE SCHEID

PhD student | CMAP - École polytechnique

After studying physics at ENS Paris-Saclay, Marie Scheid has been preparing a PhD in mathematics at CMAP (École Polytechnique) since September 2025, under the supervision of Stefano De Marco, Charles-Albert Lehalle, and Alain Durmus. Her research focuses on the generation of financial time series using generative models inspired by diffusion and flow matching.

She also aims to help structure the field through the definition of benchmarks, reference datasets, and evaluation metrics for financial data generation.

ÉCOLE POLYTECHNIQUE | INSTITUT POLYTECHNIQUE DE PARIS

GAËTAN CAILLAUT

Researcher | Dragon LLM

Gaëtan Caillaut holds a PhD in machine learning and has been working in research and artificial intelligence for nearly ten years. He specialized in natural language processing (NLP) with the emergence of BERT models.

His main focus is now on the large-scale training of Large Language Models (LLMs) in order to adapt them to industry use cases.

DRAGON LLM

ROMUALD ELIE

Researcher | Google DeepMind

Romuald Elie is an AI researcher at Google DeepMind. He holds a master's degree in mathematics and economics from Columbia University and has expertise in mathematics, economics, engineering, computer science, and artificial intelligence.
Since 2024, he has been leading a team of AI researchers and engineers based in Paris, Zurich, and Berlin, with a focus on AI alignment, diffusion models, multi-agent reinforcement learning, and their scientific applications. Previously, he was a full professor of applied mathematics at Université Gustave Eiffel (2013–2020), where he has remained affiliated since 2020, conducting research and teaching in AI, statistics, finance, and actuarial science.

DeepMind

GUILLAUME RIGAUD

Data Scientist | Ardian

Guillaume joined Ardian in September 2022 as an Analyst & Data Scientist within the Infrastructure team in Paris. He is a member of Ardian's Data Science team for the Infrastructure business.
Before joining Ardian, he spent one year as a Data Scientist at AXA Investment Managers in the Global Risk Management department, and one year as a Data Scientist at AXA France in the Digital Transformation department. Earlier in his career, Guillaume also completed an internship as an actuary and data scientist at AXA Group Risk Management.

ARDIAN

PASCAL OSWALD

Head of Market & Counterparty Risk Modelling | Natixis

Pascal Oswald is an expert in financial risk modelling with over 20 years of experience in capital markets finance. He holds a PhD in Physics (CEA, 1998–2001) and has developed his expertise in financial engineering at leading French institutions, including Société Générale, Calyon, Allianz Global Investors, Dexia, and, since 2017, Natixis CIB. Currently leading the Market and Counterparty Risk Modelling division, he oversees methodologies for Pillar 1 and Pillar 2, as well as stress-testing models. His scope also covers valuation risk methodologies, including reserves, IPV (Independent Price Verification), prudent valuation, and IFRS 9. Passionate about innovation, he is particularly interested in integrating AI and quantum computing into finance to enhance risk measurement and the pricing of derivative instruments.

NATIXIS CORPORATE AND INVESTMENT BANKING

ALAIN DURMUS

Full professor| École Polytechnique

Alain Durmus is a full professor of statistics at Ecole Polytechnique and a member of the Centre de Mathématiques Appliquées (CMAP). He earned his PhD in applied mathematics from Université Paris-Saclay in 2016. His research focuses on computational statistics, particularly developing and analyzing Monte Carlo methods, generative models, stochastic approximation, and Bayesian inference techniques. He has contributed significantly to the theoretical understanding of Markov Chain Monte Carlo algorithms,including Langevin and Hamiltonian Monte Carlo methods, as well as popular diffusion models. His work extends to applications in machine learning and inverse problems. In recognition of his contributions, Durmus received the Shiing-Shen Chern Young Faculty Award in 2023. His publications appear in leading journals and conferences in statistics and machine learning.

ÉCOLE POLYTECHNIQUE | INSTITUT POLYTECHNIQUE DE PARIS