
“Ich Kann’s Nicht Fassen, Nicht Glauben”

Markets’ Datafication and SupTech

with a Focus on Market Abuse

Filippo Annunziata

Full Professor of Financial Markets Law, Bocconi University, Milan

Visiting Professor, Ca’ Foscari, Venice · Academic Board Fellow, European Banking Institute

“I cannot grasp it, cannot believe it.”

Schumann, Frauenliebe und Leben (1840)

Schumann’s verse gives Romantic expression to the limits of comprehension when perception exceeds the mind’s capacity to order it. In today’s digital environment, supervisors face a comparable bewilderment: information itself now outstrips human understanding, and the very tools designed to enhance cognition risk outstripping it.



Datafication

Reportable information grows exponentially; analytical capacity does not keep pace.



Bewilderment

Structures we created yet cannot fully grasp — the epistemological unease of the age.



SupTech

A technological and epistemological response: translating data abundance into regulatory knowledge.

RegTech and SupTech: two responses to complexity



RegTech

REGULATED FIRMS

- Technology to enhance compliance and risk processes
- Matches regulated entities' data to regulators' information flows
- Long-standing private-sector automation
- Industry-led and ahead of the curve



SupTech

SUPERVISORS

- Digital tools that empower supervisors directly
- Refines data intelligence and analytical capability
- A recent, fast-moving debate
- Public bodies as innovators, not only enforcers

From rule-based oversight to data-driven governance

OECD: digital tools used by regulators to improve oversight, surveillance and analytical capability — with benefits for financial stability and market integrity. Its evolution unfolds across four generations.



Most supervisors today operate between the diagnostic and predictive stages.

EU financial law evolves into data governance

After the Global Financial Crisis, EU institutions realised transparency and data availability were insufficient without equally advanced analytical capacity. Supervisors faced a paradox — reportable information surged while the ability to process and interpret it stayed bounded by human constraints.

SupTech emerged to bridge this asymmetry between informational supply and supervisory absorption.



High-frequency trading

Markets process billions of orders daily, often in microseconds. Abuse can occur and vanish within fractions of a second — manual monitoring is impossible.

A DUAL LOGIC

Operational necessity

Managing extraordinary data volumes with accuracy and efficiency.

Conceptual realignment

From static thresholds to recognising evolving patterns in multidimensional data.

Two worlds collide — or combine?

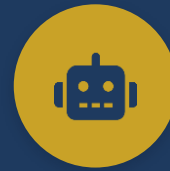


Legal-regulatory logic

Market integrity · fairness · investor protection (MAR, MiFID II)

- Clarity and determinacy
- Procedural fairness
- Declarative rules in natural language
- Demonstrable causality & reproducibility

VS



Computational logic

Pattern recognition · network analysis · predictive modelling

- Probabilistic inference
- Opacity (the “black box”)
- Continuous, adaptive learning
- Correlation & inductive reasoning

Supervised vs unsupervised learning



Supervised learning

Trained on labelled datasets from confirmed cases of abuse.

- ✓ Legal coherence — learns from conduct already unlawful
- ✓ Strong on known patterns of insider dealing, spoofing

Limit: Inherently retrospective — may miss novel manipulative tactics.



Unsupervised learning

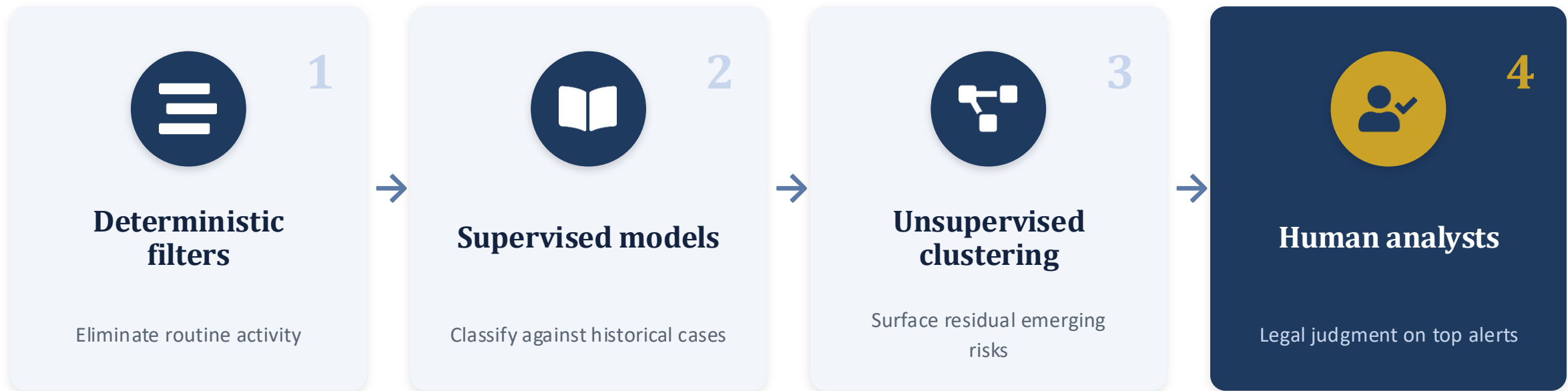
No pre-labelled data; explores records for anomalies, clusters, outliers.

- ✓ Discovers previously unrecognised behaviours
- ✓ Revealed hidden cross-venue spoofing circles

Limit: Limited interpretability — can flag, but not explain in legal terms.

The hybrid, layered surveillance architecture

Raw transaction data flow through successive layers — efficiency and normative alignment combined.



AI reduces noise and identifies patterns; humans ensure interpretive coherence.

Adaptability vs accountability

Continuous learning lets models evolve with the market — but legal certainty requires stability that can be fixed and audited at a point in time.



Algorithmic bias

Skewed training data reinforces asymmetries or over-flags certain venues.



System gaming

Participants infer the logic and adapt to evade while staying formally compliant.



Human-in-the-loop

Liability rests with institutions, not algorithms; “machine accountability” is untenable.



Controlled cycles

Model updates within sandbox testing and supervisory approval.

Adaptive supervision is inseparable from robust accountability structures.

Data quality and the translation gap

Even the most advanced algorithm cannot compensate for incomplete or inconsistent inputs. SupTech effectiveness is only as strong as its informational foundation — hence ESMA's drive for consolidated tapes and harmonised reporting.

“Data is key to understanding markets, their features and risks — and critical to effective supervision at national and European level.”

— ESMA

THE TRANSLATION GAP

“Misleading signals”



Abnormal order-to-trade ratio?

“Artificial price levels”



Sudden volume spikes?

“Abnormal volumes”



Cross-venue arbitrage pattern?

Each operationalisation embodies a legal interpretation — and a potential source of contestation.

Institutional governance & anticipatory regulation

Innovation hubs and sandboxes shift supervision from command-and-control to co-regulation, placing public authorities at the forefront of digital transformation. Ahern's anticipatory governance offers a structured response to regulatory lag and the Collingridge dilemma.



Regulatory foresight

Horizon scanning, scenario planning, stress-testing multiple futures.



Regulatory experimentation

Sandboxes, pilot regimes, policy labs, time-limited exemptions.



Regulatory learning

Institutionalised feedback loops and ex-post evaluation.

The EU — MiCAR, the AI Act, the DLT Pilot, Interoperable Europe sandboxes — is emerging as a pioneer of anticipatory regulation; yet the DLT Pilot's difficulties show the complexity involved.

Market-abuse surveillance: where SupTech bites

The supervisory and sanctioning process runs in four phases. SupTech's impact concentrates on the first — defining a region of normality to capture anomalous shifts.



Market-abuse detection is a species of anomaly detection: define normality, then capture the deviations that may point to abusive practice. Automation reshapes detection — not the legal assessment that follows.

Two surveillance paradigms



Information-based

NON-TRANSACTIONAL SOURCES

- News articles, social media, public information
- NLP detects language & sentiment patterns
- Better at new or emerging risks

Hard to design: large volumes of dynamic information processed adaptively.



Transaction-based

ORDER & TRADE DATA

- Trading patterns and behaviours in market data
- ML flags insider trading / manipulation
- Strong on established forms of abuse

Most live systems today are indicator-based and real-time.

Common surveillance models for market abuse



Rule-based

Pre-defined thresholds flag unusual size, value or out-of-hours trades.



Pattern recognition

ML trained on historical data recognises anomalous behavioural patterns.



Network analysis

Social-network techniques expose insider rings and collusion clusters.



Text mining

NLP on news, posts and filings detects telling language or sentiment.

Across every model, SupTech combines supervised learning (known patterns) with unsupervised learning (outliers) to surface insider trading that traditional methods miss.

CONSOB: a model of supervisor–academia cooperation

Italian Consiglio di Stato (2019–20) admits automated decision-making under three cumulative conditions — knowability of the logic, non-exclusivity (human review), and non-discrimination — echoed by the AI Act.



Insider trading

Vast datasets, growing analytical complexity



KID / PRIIPs review

NLP & image recognition compare disclosures



Online illicit activity

LLM pairs complaints & web data (MiFID/MiCAR)



Greenwashing (Trento)

ClimateBERT / ESG-BERT on prospectuses



Human-in-the-loop throughout. Three SNS-Pisa proofs of concept (K-means; Statistically Validated Networks; PCA + autoencoders) generate alerts only — no autonomous authority. A 2024 reform created a Division for Informatics & AI, mitigating regulatory-capture risk through research-institution cooperation.

SupTech in practice across the EU

**AMF**

FRANCE

ICY

Advanced data-analytics platform monitoring trading activity to detect potential market abuse.

**BaFin**

GERMANY

Project ALMA

Integrated automated alarm and market-monitoring system for insider trading and manipulation.

**AFM**

NETHERLANDS

STOR feedback loop

Millions of transactions/month; microstructure indicators; missed cases fed back to refine detection — a self-learning ecosystem.

Global benchmarks in supervisory technology



SEC · MiDAS

UNITED STATES

Born after the 2010 Flash Crash. Processes complete order-book data from all US equity exchanges — billions of records daily — reconstructing events at microsecond precision to detect spoofing, layering and quote stuffing.

Draws on CAT and SIP feeds; sits alongside NEAT and ATLAS — a benchmark for supervisory innovation worldwide.



CSA · MAP

CANADA

Post-trade analytics across fragmented jurisdictions (Kx Systems). Cross-market, cross-venue detection of spoofing, layering and collusion.



ASIC · MAI

AUSTRALIA

Operating 10+ years. Real-time, cross-market analytics; trials of ML, social-media analytics and web-crawling against predatory trading.

Expressed and unexpressed risks



Cybersecurity

Bugs, hacks and IT failures whose interruption can ripple through markets.



De-skilling

Over-reliance breeds complacency and erodes human expertise.



Fragmentation

Siloed databases across institutions cause coordination failures.



Algorithmic bias

Skewed data and opaque models threaten fairness and due process.



Black box & XAI

Explainability often trades off against accuracy — a governance dilemma.

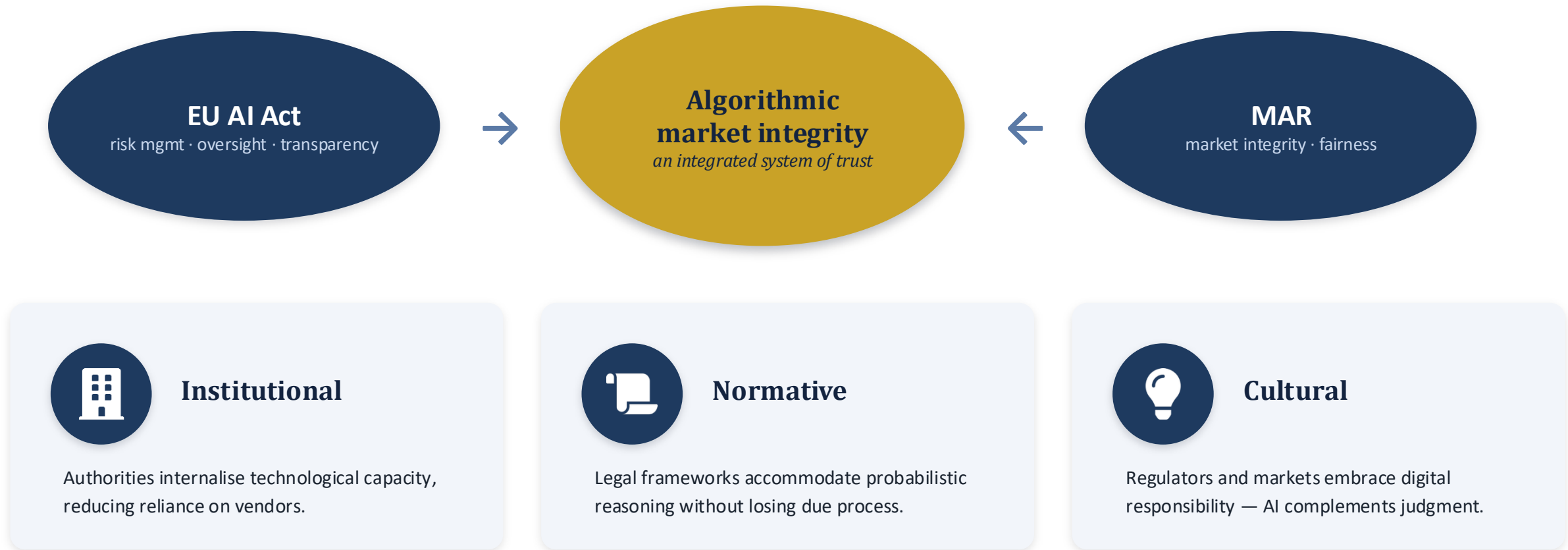


GDPR / cross-border

Purpose limitation constrains repurposing data for AI training.

Real-time supervision blurs the line between supervision and adjudication — procedural safeguards must keep human oversight meaningful.

EU AI Act + MAR: algorithmic market integrity



CONCLUDING REMARKS

“I cannot grasp it, cannot believe it.”

*SupTech stands as both an instrument and a mirror —
extending regulatory reason into the digital domain, and
reflecting the fragility of that reason when mediated by code.*



Embed in formal governance

Explicit acts defining purposes, data sources, safeguards — full ex-post reviewability.



Layered architectures

Rule-based screening, ML, and structured human oversight, without displacing judgment.



Data quality as imperative

A regulatory question, not an ancillary technical one.



Institutional capacity

Internalise interdisciplinary expertise; converge on EU governance standards.

Its enduring test lies not in how much it can know, but in how wisely it can judge.