

September 15, 2026

Submitted via email to pubcom@finra.org

Jennifer Piorko Mitchell
Office of the Corporate Secretary
FINRA
1700 K Street, NW
Washington, DC 20006

Re: Regulatory Notice 26-15 — Request for Comment on Modernizing FINRA’s Best Execution Guidance

Dear Ms. Mitchell:

Quantex Technologies Inc. appreciates the opportunity to comment on Regulatory Notice 26-15. We limit our response to three of the questions raised in the Notice, in areas where we believe we can offer something beyond restatement of the existing framework. We take no position on the merits of the Commission’s proposal to rescind Rule 611; our comments assume rescission solely for the purpose of addressing the questions FINRA has posed.

This letter is submitted by Carlos R. Cabana, Chief Executive Officer and Founder of Quantex Technologies Inc. The perspective it reflects is one of a practitioner. I have spent more than 30 years in institutional capital markets, principally in equities — at Knight Capital, Cantor Fitzgerald, Topeka Capital, and CastleOak Securities — across trading, operations, and supervisory functions, and earlier in foreign exchange at Morgan Stanley and Bear Stearns. My registration history under CRD# 2243792 includes FINRA Series 7, 24, 63, 55, and 3. The Series 24 is the most relevant of those to this Notice: the questions FINRA raises here are supervisory questions before they are technological ones. Quantex Technologies Inc. develops governance and evidence-capture technology for automated decision systems; that commercial interest is disclosed in the note below.

I note the foreign exchange experience deliberately. FX has never operated under a trade-through rule, and best execution in that market has always been a matter of demonstrable judgment rather than comparison to a protected quote. Much of what this Notice contemplates is familiar terrain there.

I write in my personal capacity and on behalf of Quantex Technologies Inc. I do not write on behalf of, and these views should not be attributed to, any broker-dealer or other firm with which I am or have been associated. Observations below regarding institutional order handling reflect general industry practice as I have encountered it across multiple firms over three decades; they do not describe the practices of any particular firm.

We have tried throughout to write a letter that would be equally useful to a firm that never purchases a system from anyone.

I. Summary of Position

Rule 611 has functioned for two decades as more than a routing constraint. It has functioned as an **objective external referent** — a publicly observable fact against which a firm’s conduct could be compared without inquiry into the firm’s reasoning. A trade-through either occurred or it did not.

Best execution under Rule 5310 has never depended on that referent, and FINRA is correct that the reasonable diligence standard applies with or without it. But the *examinability* of best execution has quietly depended on it. Where an outcome could be measured against the protected quote, the sufficiency of a firm’s reasoning was rarely the operative question. Absent Rule 611, it becomes the only question.

Reasonableness of judgment cannot be examined directly. It can be examined only through the record of how the judgment was formed. This means that the firm’s compliance problem and FINRA’s oversight problem — Questions 9 and 10 — are not two problems but one:

What constitutes a sufficient record of a routing or handling decision?

We propose that FINRA answer this question in interpretive guidance rather than rule text. We agree with FINRA’s preliminary view that Rule 5310 and its supplementary material remain appropriately calibrated, and we do not propose amending them. What is missing is not a standard of conduct but a **standard of evidence**.

II. The Evidentiary Sufficiency Framework

A. The relationship to existing audit trail infrastructure

We should address at the outset the obvious question of whether the industry already has what we are describing.

The Consolidated Audit Trail captures the lifecycle of an order comprehensively — receipt, modification, routing, execution, allocation — across reporting members. It is an event record, and an excellent one. But it is a record of what happened *to* an order, and what rescission of Rule 611 puts in issue is why a firm did what it did.

Three things material to the reasonable diligence inquiry fall outside its scope. It does not capture the market state the firm actually observed at the moment of decision, which can differ from consolidated data in exactly the circumstances where execution quality is later contested. It does not capture the alternatives within the firm’s opportunity set that were considered and rejected — under Rule 611 the significant alternative was defined externally by the protected quote, and absent the rule nothing defines it but the firm’s own record. And it does not capture which version of the firm’s routing logic, policy, or model governed the decision.

The distinction we draw is therefore narrow and, we think, precise: **the existing audit trail records events; what rescission exposes is the absence of a record of reasons**. A substantial portion of what we describe below can be constructed on infrastructure members already maintain. That bears materially on cost, and we return to it.

B. The five elements

We propose that guidance identify five elements of a sufficient decision record. A record satisfying all five permits a reviewer — supervisor, examiner, or adjudicator — to reconstruct the decision without relying on the firm's later characterization of it.

1. Observed state. The market data actually available to the deciding system or person at the moment of decision — not the state as later reconstructed from consolidated sources. These diverge, sometimes materially, and the diligence inquiry concerns what the firm knew, not what was knowable.

2. Available alternatives. The venues, handling paths, and execution methods within the firm's opportunity set at that moment, including those not selected. Rule 611 made the counterfactual implicit: a firm was required to account for the protected quote whether or not it routed there. Without the rule, the alternatives considered — and the basis for excluding them — must be affirmatively recorded, or the firm's diligence cannot be distinguished from its default configuration.

3. Governing logic. The specific policy, routing logic, model, or supervisory instruction in force at that moment, identified with sufficient specificity that the version applied is unambiguous. A firm that cannot establish which version of its logic governed a decision cannot establish that the decision was governed at all.

4. Decision linkage. The connection between elements 1 through 3 and the outcome — how the observed state, evaluated against the alternatives, under the governing logic, produced the handling that occurred.

5. Record integrity. Evidence that the record was created contemporaneously with the decision and has not been altered since. A reconstruction assembled after an inquiry begins is an assertion about a decision; it is not a record of one. This distinction is the difference between a defensible file and a persuasive one.

C. Proportionality and cost

We want to be direct about the burden these elements impose, because we think the candid account is more useful to FINRA than a reassuring one.

The elements are technology-neutral as specified — none names a technique or a product. But capture cost is not flat. A firm handling a modest volume of orders with substantial human involvement can satisfy all five through disciplined logging, versioned written procedures, and periodic integrity verification. A firm routing at high volume through automated systems cannot, and will require systems to do it. We do not think this should be obscured.

We also do not think it is unfair. The reconstruction burden grows with the number of decisions a firm makes without direct human attention, which is a burden the firm creates by choosing to automate. A framework that imposed identical cost regardless of automation would subsidize the firms whose decisions are hardest to review.

On magnitude, we would offer two observations that we believe reduce the estimated burden materially:

Capture and retention should be separated. Capture must be complete to be useful, since a record assembled selectively is not contemporaneous. Retention need not be. We suggest guidance contemplate tiered retention — full decision records for a defined rolling window sufficient for supervisory review and examination cycles, with structured summaries preserving the elements in compressed form thereafter. Retention volume, not capture, drives most of the cost in comparable regimes.

The marginal data is smaller than it appears. Order lifecycle events are already captured. Market data is already ingested and, at many firms, already recorded. What is typically absent is the linkage between them, the identification of the governing logic version, and the record of rejected alternatives. For most members this is an integration and instrumentation problem rather than net-new data collection, and the cost profile of the two is very different.

We would encourage FINRA to seek quantitative estimates from members of varying size before settling any guidance, and we would note that a framework which proves unaffordable for smaller members will be satisfied nominally rather than substantively — which is a worse outcome than a less ambitious standard honestly met.

III. Response to Question 9 — Safe Harbors

FINRA asks whether one or more Rule 5310 safe harbors should be developed, and identifies executions pursuant to documented regular and rigorous methodologies as a candidate.

We support a safe harbor of that character, with three qualifications.

It should attach to the record, not the outcome. A safe harbor keyed to a benchmark result would reintroduce, by another route, the outcome-comparison logic that rescission of Rule 611 removes — and would do so using benchmarks (NBBO, VWAP, TWAP) whose reliability in the post-rescission environment is precisely what the Notice questions elsewhere. A safe harbor keyed to evidentiary sufficiency does not depend on any benchmark surviving.

It should be a rebuttable presumption, not immunity. A firm whose decision records satisfy the five elements, and whose regular and rigorous reviews are documented against those records, should receive a presumption that reasonable diligence was exercised. Absolute safe harbor treatment would invite the construction of methodologies designed for their protective effect rather than for execution quality, and would be difficult to defend to investors.

We suggest the presumption be rebuttable on any of four grounds: bad faith; undisclosed conflict; systematic disregard of the firm's own documented methodology; or **persistent outcome disadvantage to an identifiable class of orders or customers, whether or not any individual decision departed from the methodology.**

We include the fourth deliberately, and we recognize it is the one that makes the safe harbor meaningfully harder to rely on. A methodology can be faithfully documented, faithfully followed, and systematically bad. Confining rebuttal to process failures would create precisely the shield that critics of documentation-based safe harbors reasonably fear, and would be inconsistent with our own position in Section VI.D that order-level and aggregate review detect different failures. A safe harbor that cannot

be pierced by outcome evidence is not a presumption about diligence; it is a conversion of the duty into a paperwork obligation.

It should not excuse the review obligation. The presumption should be available only where the documented methodology is itself subject to periodic reassessment against competing venues, consistent with Supplementary Material .09(b). A methodology that is faithfully followed but never re-examined is a formalized failure to review.

IV. Response to Question 10 — FINRA Oversight and Data

FINRA asks how it should approach oversight absent a trade-through rule, and what data it should consider.

We suggest that the answer involves a shift in the *object* of examination rather than an expansion in the *volume* of data collected.

Existing execution quality disclosure regimes are outcome-oriented. They are valuable and should continue. But outcome data answers the question “what happened,” and the post-rescission inquiry is “was the diligence reasonable” — a question outcome data can raise but cannot resolve. Expanding outcome collection would also compound the benchmark problem: statistics computed against a reference price that no longer carries regulatory significance measure conformance to a convention rather than quality of judgment.

We suggest instead that examination focus on **reconstruction capability**. The examinable question becomes: *can this firm, for a sampled order, produce a contemporaneous record satisfying the five elements?* This approach has several properties we believe FINRA would find useful:

- It is answerable objectively. Either the record exists and reconstructs, or it does not. It does not require FINRA to second-guess a trading decision in order to identify a deficiency.
- It is diagnostic rather than merely punitive. A firm that cannot reconstruct its decisions has a supervisory failure under Rule 3110 that is worth surfacing on its own terms, independent of whether any customer was disadvantaged.
- It scales to automation. As order handling moves further from direct human attention — including into systems using machine learning or generative models — reconstruction capability remains meaningful where retrospective human explanation does not.
- It preserves the principles-based standard FINRA wishes to retain. It specifies what a firm must be able to *show*, not what a firm must *decide*.

We would suggest that any such examination approach be introduced through guidance and examination priorities rather than new reporting obligations, at least initially, and that FINRA consider publishing anonymized observations on reconstruction adequacy across firm types — which would give members a concrete calibration point that no amount of principles-based drafting can supply.

V. Response to Section 9, Question 2 — Artificial Intelligence in Order Handling

FINRA asks whether members currently use AI in order handling, routing, or execution decisions, whether they plan to, and whether AI-specific best execution guidance is warranted.

We offer no comment on Question 1 regarding tokenized securities, which other commenters are better positioned to address.

A. Current practice, as we observe it

Our observation is that automated statistical decision-making has been embedded in institutional order handling for well over a decade, and that the recent introduction of large language models and other generative systems has followed a distinct pattern: adoption has been fastest in functions adjacent to the order path rather than within it.

Generative systems are today most commonly found in surveillance triage, exception review, drafting of internal review memoranda, research summarization, and client communication. Machine learning of the non-generative variety is considerably more embedded in the order path itself — in venue selection and scheduling, liquidity and fill-probability estimation, market impact modeling, and the parameterization of execution algorithms.

That boundary is eroding, and we do not believe it will hold. The relevant question for FINRA is therefore not whether members use AI in order handling — some do, more will — but whether the existing evidentiary assumptions of Rule 5310 guidance survive when the deciding system is non-deterministic. We would encourage FINRA to survey members directly on current usage, as we suspect the answers vary widely by firm type and are not well captured by any public source.

B. “AI” is not a regulatory category

We would urge FINRA to resist drafting guidance keyed to the term “artificial intelligence.” The term now spans techniques whose supervisory characteristics differ so substantially that a single standard would be simultaneously overbroad and under-inclusive. A firm that has run a statistically optimized smart order router since 2012 is doing something the existing guidance already addresses; a firm routing on the output of a generative model is not.

We suggest guidance instead distinguish systems by three characteristics that bear directly on evidentiary sufficiency:

Determinism. Does the system, given identical inputs and an identical version, produce an identical output? Deterministic systems can be replayed. Non-deterministic systems cannot, and for those the contemporaneous record is the only record that will ever exist.

Enumerability. Can the operative logic be stated in advance, in full, in a form a supervisor can read? Rule-based logic can. Parametric models can be described but not enumerated. Generative systems can be neither.

Stability between supervisory events. Does the logic change only at a controlled release, or can it change continuously? A system that adapts online breaks the assumption underlying most supervisory

frameworks: that a firm's written procedures describe a stable artifact between reviews. Where a model updates continuously, the firm's effective procedures are changing without any event that triggers review.

These characteristics can be assessed by a compliance officer without technical training, and they will remain meaningful as the underlying technology changes — which a definition keyed to today's architectures will not.

C. The central problem: post-hoc explanation is not evidence

We regard this as the most consequential point in our letter.

When a trading decision is questioned, the traditional supervisory method is to ask the decision-maker to explain the reasoning. This method has an implicit premise: that the explanation bears a causal relationship to the process that produced the decision. For a human trader the premise is imperfect but broadly serviceable.

For a generative model asked to explain its own prior output, the premise does not hold. Such a model will produce a fluent, internally coherent, and often persuasive rationale that need not correspond to the computation that produced the output. The explanation is generated, not retrieved. It may be accurate; nothing in the mechanism ensures that it is, and its persuasiveness carries no information about its accuracy.

We want to state the scope of this claim carefully, because we do not think it is true of automated systems generally. For parametric and tree-based models, established attribution methods can identify which inputs drove a given output, and model risk management practice for such systems is reasonably mature. Firms using those techniques in the order path are not without explanatory tools, and guidance should not suggest otherwise. Our claim is narrower and, we think, harder to work around: **it concerns a system's narration of its own prior reasoning, and it is at its strongest precisely where the technology is newest and adoption is accelerating fastest.**

This has a direct consequence for best execution supervision. A firm that captures only the decision and the ability to interrogate the model later has captured nothing of evidentiary value — it has acquired the ability to manufacture a justification for any outcome, which is worse than having no record at all, because the resulting file will look complete.

The framework we set out in Section II is therefore not merely useful in this context but necessary. Where the deciding system cannot be meaningfully asked why, the observed inputs, the identified model version, the alternatives available, and the recorded output must be captured at the time of decision or the decision becomes permanently unreviewable. We would encourage FINRA to state plainly in guidance that a model-generated explanation of a model's own prior output does not, standing alone, satisfy any supervisory or recordkeeping expectation.

D. Human review and the limits of the “human in the loop”

We anticipate that many members will describe human review or override capability as the control that addresses these concerns. We would encourage FINRA to examine such claims carefully.

A human who approves a recommendation they cannot independently derive, within a timeframe that does not permit independent derivation, is not exercising supervision. The literature on automation bias is consistent: reviewers presented with a confident machine recommendation approve at very high rates, and the rate rises as time pressure and volume rise. In an order-handling context both are high by construction.

We suggest guidance provide that a human approval step constitutes meaningful supervision only where the reviewer has (i) access to the decision record contemporaneously, (ii) sufficient time and information to reach an independent view, and (iii) practical authority to override without adverse consequence. Absent those conditions, the human step should be understood as documentation of the machine's decision rather than review of it — and firms should not receive supervisory credit for it. We think this distinction matters more than any technical control, because it is where the appearance of accountability and the substance of it diverge most cleanly.

E. Third-party model dependency

A related gap deserves attention. Where a firm's order handling depends on a model supplied by a vendor, the firm's effective routing logic can change when the vendor updates the model — potentially without notice, without a version identifier the firm can record, and without any event that triggers the firm's own change management process.

This is a live problem today and it will grow. FINRA has long held that a firm cannot transfer its best execution obligations to another party. We suggest guidance make explicit that this principle extends to model provenance: a firm relying on a third-party model in the order path should be able to identify which version governed a given decision and should have contractual or technical means of learning when that version changes. A firm that cannot do so has, in substance, delegated its routing logic to a party outside its supervisory reach.

We note that this recommendation constrains vendors, including firms situated as we are. We think it is the correct standard notwithstanding.

F. Recommendation

We do not recommend an AI-specific best execution standard. The duty under Rule 5310 is unchanged: reasonable diligence to ascertain the best market, assessed on facts and circumstances. Creating a parallel standard would invite arbitrage over which systems fall within it and would date quickly.

We recommend instead that FINRA address AI within guidance on evidentiary sufficiency — specifying what a firm must be able to demonstrate when the deciding system is non-deterministic, non-enumerable, or unstable between supervisory events. Guidance framed that way requires no forecast about which technologies members will adopt, and it applies with equal force to systems that do not yet exist.

VI. Response to Section 3, Question 3 — Order-by-Order Review and Automated Routing Logic

FINRA asks what processes an order-by-order review should involve, how it differs from a regular and rigorous review, whether a smart order router with built-in best execution logic should constitute an order-by-order review, and whether aggregate periodic review should be required as a complement or alternative.

A. A router is a decision-maker, not a reviewer

We would answer the central question in the negative, and we think the reason is worth stating precisely because the contrary view is superficially attractive.

A smart order router with embedded best execution logic is the mechanism by which the routing decision is *made*. Review is the evaluation of a decision by something other than the process that produced it. To treat the router as satisfying the review obligation is to collapse the actor and the evaluator into a single object and call the result oversight.

The analogy in human terms is exact: a trader who considers execution quality carefully before each order has not thereby conducted an order-by-order review of their own trading. The care is genuinely valuable. It is not review, and no supervisory framework has ever treated it as such. Nothing about implementing the same care in software changes the analysis.

This matters most where the requirement matters most. FINRA has required order-by-order analysis for internalized flow because of the conflict inherent in a firm executing against its own customer from inventory. That requirement is a conflict control. A conflict control that is administered by the conflicted process itself is not a control.

B. What automated routing does provide: reviewability at scale

We do not want the answer to end there, because the question contains a real insight that deserves to be captured.

The historical difficulty with order-by-order review has never been conceptual. It has been that human examination of individual orders does not scale, which is why Supplementary Material .09 offers regular and rigorous review as the alternative in the first place. A properly instrumented router changes that constraint. It produces, for every order, a record of the kind described in Section II — what was observed, what alternatives existed, what logic governed, what resulted.

That is not review. It is the precondition for review, and it is a significant one: it makes genuine order-by-order evaluation feasible across the whole population rather than a sample, for the first time since the requirement was written.

So the correct answer to FINRA's question, in our view, is that the router should not be deemed to constitute order-by-order review, but that the presence of a router producing sufficient decision records should be recognized as what makes automated order-by-order review possible — and automated review, properly constituted, should be recognized as satisfying the requirement.

C. The condition we propose: independence of the evaluating logic

We suggest that guidance permit automated order-by-order review where the evaluating logic is independent of the deciding logic in four specific respects:

Separately specified. The criteria by which an execution is evaluated are written down as their own artifact, not inferred from the router's configuration.

Separately versioned. A change to the routing logic does not silently change the evaluation criteria, and the firm can identify which version of each governed any given order.

Not co-optimized. The evaluation does not use the same objective function the router optimizes. A router optimized for fill probability that is evaluated on fill probability will report success reliably and tell the firm nothing. Where the deciding and evaluating logic share an objective, the review is tautological.

We anticipate the fair objection that every evaluation criterion is some objective function, and that this condition therefore forbids nothing coherent. The answer is that the criterion should be derived from the customer's expectations for the order and from realized outcomes, rather than from the router's internal optimization target. An algorithm optimizing schedule adherence should be evaluated on realized cost against arrival, not on schedule adherence. The distinction is not that the evaluation lacks an objective; it is that the objective is drawn from what the customer was owed rather than from what the system was built to maximize.

Escalating to a human. The evaluation is capable of producing an exception that a person actually sees, within a timeframe in which something can be done about it, with the conditions set out in Section V.D applying to that person's review.

Firms will observe that this is more demanding than current practice, and it is. We would note that it is considerably less demanding than the human order-by-order review the rule contemplates on its face, which is why we think it represents a workable modernization rather than a tightening.

D. Aggregate review should be a complement, not an alternative

We support explicitly requiring aggregate periodic review alongside order-by-order review rather than in place of it, because the two detect different failures.

Order-level review detects the individual decision that was wrong — the conflicted internalization, the ignored better price. Aggregate review detects the failure in which every individual decision passes its own test and the distribution is nonetheless bad: gradual drift in venue mix, systematic disadvantage to a subset of orders or customers, a routing table that has quietly stopped reflecting current execution quality. A firm can hold a clean order-level record for every order in a quarter and still have routed poorly all quarter. Neither review substitutes for the other, and a firm permitted to choose will predictably choose the one that produces fewer exceptions.

E. A note on institutional order structure — Section 4, Questions 5 and 8

We add a brief observation on two questions raised elsewhere in the Notice, because they bear directly on what unit of activity is properly reviewable.

FINRA asks whether price-oriented standards adequately capture pre-negotiated block trades, and how best execution should apply to parent orders worked as multiple child orders. In our experience these are the two settings where order-level, price-anchored review fits worst, for a shared reason: in both, the reviewable event is not the event the review examines.

For a block negotiated over an extended period, the print is the last and least informative moment in the sequence. The judgments that determined the quality of the outcome — which counterparties to approach and in what order, how much to disclose and to whom, how to size and stage the inquiry, and repeatedly whether to transact at all — occurred earlier and mostly consist of decisions not to trade. A review that measures the execution against the market at the moment of the print is measuring the residue of the process rather than the process.

For parent orders, evaluating each child fill against contemporaneous market conditions can produce excellent statistics for an execution that was poor precisely because of the child pattern. Information leakage created by the schedule is invisible at child level and shows up only in the parent's realized cost against its own arrival.

We therefore suggest that guidance identify the parent order as the reviewable unit where an order is worked, with child-level records serving as supporting evidence, and apply the elements in Section II at parent level — what was known at each decision point, what alternatives existed, what schedule or negotiation logic governed, and why deviations occurred. This is more consistent with how these orders are actually managed, and it makes the review answerable rather than merely computable.

We acknowledge a tension with our position elsewhere in this letter that Rule 5310 requires no textual amendment. The rule applies to any transaction, and treating the parent as the reviewable unit could be read as narrowing that. We do not think it does: the duty continues to attach to every transaction, and what we propose concerns the level at which reasonable diligence is *assessed* rather than the transactions to which it applies. A child fill executed in disregard of the customer's interest would remain a violation on its own terms. If FINRA concludes that the point cannot be reached interpretively, we would prefer it be addressed in guidance as an application of the existing standard rather than left unresolved, and we flag the question rather than assume our answer to it.

VII. Conclusion

The three areas we have addressed converge on a single proposition. Rescission of Rule 611 does not change the duty of best execution, but it removes the external referent that has made the duty examinable without inquiry into a firm's reasoning. What replaces that referent cannot be another benchmark, because the reliability of benchmarks in the post-rescission environment is itself uncertain. What can replace it is a standard governing what a firm must be able to demonstrate about how a decision was made.

Such a standard preserves the principles-based framework FINRA wishes to retain, because it specifies what a firm must be able to show rather than what a firm must decide. It remains meaningful as order handling becomes more automated, including in settings where retrospective explanation is unavailable or unreliable. And it gives examiners an objective question to ask that does not require second-guessing a trading judgment.

We appreciate FINRA's consideration and would welcome the opportunity to discuss any of these points further.

Respectfully submitted,

Carlos R. Cabana

Chief Executive Officer & Founder

Quantex Technologies Inc.

FINRA CRD# 2243792 | Series 7, 24, 63, 55, 3

Quantex Technologies Inc.

1500 Stony Brook Road, Stony Brook, NY 11794

Office: (631) 246-0861

ccabana@quantex-tech.com

www.quantex-tech.com

Disclosure of Commercial Interest

Quantex Technologies Inc. develops governance and evidence-capture technology for automated decision systems, including in regulated financial contexts. Recommendations in this letter that would create demand for such technology are identified as such, and we have sought throughout to propose standards that can be satisfied without purchasing any system, including our own.