March 14, 2024

Christopher Kirkpatrick, Secretary
Commodity Futures Trading Commission
Three Lafayette Centre
1155 21st Street, NW
Washington, DC 20581

Re: Request for Comment on the Use of Artificial Intelligence in CFTC Regulated Markets

Dear Mr. Kirkpatrick:

Thank you for the opportunity to comment. It's refreshing to see a market regulator start a discussion about AI technology. The industry and the public can compliment or criticize AI and the government can evaluate these technologies after a full public hearing. A different market regulator had that chance last year and didn't bother. Rather than talk to the industry, issuers, and the public about AI technology, last year the SEC took a different path and simply rubber-stamped a large stock exchange's plan for an AI order type, a first of its kind.[1] To approve the order type, and without notice, the SEC also gutted a bedrock reform it adopted 25 years ago and replaced it with a new standard.[2] The SEC's new standard redefines what a securities exchange can do and is so vague almost any exchange rule, policy, or procedure complies. As I discuss in more detail later, exchanges can meet the SEC's standard with a Ouija board or a monkey throwing darts. It's pretty funny to hear the SEC preach about AI and market integrity these days.[3] When it had the chance to do the work, the SEC walked out.

My comments here will be about AI in market center trading facilities, and specifically about the public markets using AI *deep learning* technology to handle participant orders. That's what the SEC rubber-stamped last year, throwing out decades of precedent to cap off a brawl in a corner of the SEC's rule filing processes.[4]

Market center trading systems today are predictable and knowable. Give them a set of inputs and they will do the same thing every time and anyone can prove it. Developers dictate specific results according to market rules and then data goes in - orders, reference data, various settings, and so on - and the system produces exactly those results.[5] These rule-based systems are *deterministic* and they give market centers, participants, and regulators *certainty* and *transparency*. Everyone knows what they should do and they do it. Participants and regulators can *police* markets to make sure they follow the rules and hold them accountable.

Deep learning methods will reverse all of this. That's what kicked off the brawl, which I believe was littered with regulatory, mathematical, and statistical gibberish and crude techno-onanism. I also believe it caught the SEC completely by surprise, and unprepared and uninformed. I was a contestant in that and

---

[1] See SEC File No. SR-NASDAQ-2022-079, "Proposed Rule Change to Amend Rules 4702(b)(14) and (b)(15) Concerning Dynamic M-ELO Holding Periods." The File is available at https://www.sec.gov/rules/sro/national-securities-exchanges?aId=&sro_organization=192811&title=&release_number=&file_number=SR-NASDAQ-2022-079&year=All.

[2] See letter from Edgar T. Snodgrass dated September 14, 2023 (satirical), commenting on SEC File No. S7-02-22, Release 34-94062, "Amendments Regarding the Definition of 'Exchange' and Alternative Trading Systems (ATSs) That Trade U.S. Treasury and Agency Securities, National Market System (NMS) Stocks, and Other Securities." Available at https://www.sec.gov/comments/s7-02-22/s70222-258759-606783.pdf.

[3] See for example "Gary Gensler urges regulators to tame AI risks to financial stability," *Financial Times*, October 14, 2023.

[4] SEC Release 34-98321 and exhibits and letter from Joseph Saluzzi, Themis LLC, dated Janary 25, 2023 ("Themis"), letter from R. T. Leuchtkafer dated January 31, 2023 ("Leuchtkafer 1"), letter from R.T. Leuchtkafer dated May 2, 2023 ("Leuchtkafer 2"), letter from R. T. Leuchtkafer dated May 30, 2023 ("Leuchtkafer 3"), letter from R. T. Leuchtkafer dated August 11, 2023 ("Leuchtkafer 4"), letter from R. T. Leuchtkafer dated September 28, 2023 ("Leuchtkafer 5"). All letters are available at https://www.sec.gov/comments/sr-nasdaq-2022-079/srnasdaq2022079.htm.

[5] There are a very few and narrow examples where public markets decide something with a random process and so aren't entirely predictable. One example is a random reserve order, where the disclosed portion of an order is randomly determined.

I'll go through my take on the matter at length below. Thankfully the CFTC has already done much better with its well-publicized request for comment.

First Things First

Deep learning is a category of AI technologies. Deep learning methods use neural networks - a complex fabric of weighted nodes assembled in a decision tree, in effect - to estimate a favored result from a set of inputs. Broadly, a developer sets an overall plan, data goes in, and after a series of analyses and trials - even thousands or millions of these - a result pops out. SEC Chair Gary Gensler co-authored a paper about deep learning in 2020, shortly before he joined the SEC in 2021. He wrote the technology includes "five intrinsic features of hyper-dimensionality, non-linearity, non-determinism, dynamism, and complexity" and "three heightened challenges of limited explainability, fairness, and robustness" as well as "an insatiable hunger for data."[6] He said these features and challenges were "a significant departure from existing technologies,"[7] they "may lead to financial system fragility,"[8] "market transparency is likely to decrease,"[9] and that current regulatory regimes "are likely to fall short."[10] At first I thought the paper was meant to help, to teach and caution about deep learning. After the SEC rubber-stamped that AI order type it dawned on me the paper was, as in a famous "Twilight Zone" episode, a cookbook.

The most troubling "significant departure" from existing market technology is that deep learning systems are not deterministic. They can be altogether unpredictable. The same data inputs can produce different results over time, even moment to moment, and the technology's specific behavior can be so complex it is unexplainable and unauditable. And while trading centers today are as objectively fair and unbiased as the rules they implement, deep learning methods can easily and silently discriminate against one group in favor of another regardless of the rules. Participants and regulators won't be able to police markets and have to trust they handle things correctly.

Deep learning can eat mountains of data. Its fitness depends on how developers select, generate, sample, characterize, and update that data. For market center trading systems some obvious data sources for these methods are participant orders and trades. To implement a deep learning method, market centers have to decide what data elements feed the algorithms, such as order type descriptors, order terms like disclosed and undisclosed prices and volumes, or trade terms like price and size. They'll choose participants and instruments to include and what participant and instrument classifiers to use. Markets will also pick time periods to include and if and how data elements are categorized and aggregated.   All these choices have unprecedented consequences.

Market centers also decide what a deep learning method will do and why. Will it set auction times or circuit breaker values, or the timing and duration of other market operations, to favor a particular outcome?[11] Will it control and adjust participant order terms like order type, displayed or undisplayed prices, displayed or undisplayed volumes, or time-in-force to affect a participant's results?[12] Will it set order priorities and select parties to a trade? These choices also have unprecedented consequences. Markets and regulators then decide what's disclosed about these methods and where and how it's disclosed. That has unprecedented consequences too.

---

[6] Gary Gensler and Lily Bailey, "Deep Learning and Financial Stability," November 1, 2020, page 3. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3723132.
[7] Gensler and Bailey, page 3.
[8] Gensler and Bailey, page 1.
[9] Gensler and Bailey, page 26.
[10] Gensler and Bailey, page 1.
[11] As with File No. SR-FINRA-2022-032, "Notice of Filing of a Proposed Rule Change Relating to Alternative Display Facility New Entrant." The File is available at https://www.sec.gov/rules/sro/finra?aId=&title=&release_number=&file_number=SR-FINRA-2022-032&year=All.
[12] As with SEC File No. SR-NASDAQ-2022-079, "Proposed Rule Change to Amend Rules 4702(b)(14) and (b)(15) Concerning Dynamic M-ELO Holding Periods." The File is available at https://www.sec.gov/rules/sro/national-securities-exchanges?aId=&sro_organization=192811&title=&release_number=&file_number=SR-NASDAQ-2022-079&year=All.

Unlike today's technology:

- Deep learning is not deterministic. Results are unpredictable, unexplainable, and unauditable;
- Deep learning depends on the data it's fed and how developers select, generate, sample, characterize, and update that data;
- Despite intentions, deep learning can discriminate against participants or participant classes;
- Market centers can't fully explain to regulators or participants specific results. Why deep learning did one thing as opposed to any other can be so complex it's unknowable;
- Because specific results are unpredictable, unexplainable, and unauditable, regulators and participants can't police what market centers do and why.

Should public markets do deep learning, and if they do, how should they do it? What follows is my view of what happened with the first AI-driven order type approved by the SEC. Along the way I'll list some questions and considerations.

Dynamic M-ELO, An AI Case Study

In 2018 Nasdaq launched a nondisplayed order type called the Midpoint Extended Life Order ("M-ELO") in its stock market. Nasdaq intended it for participants who "seek liquidity at the midpoint of the NBBO" and may have "longer term investment horizons," and we can guess it was meant to appeal to pension and mutual funds, among others.[13] The order type's novelties were that it came with a particular time-in-force term and traded only with other M-ELO orders. The time-in-force term set a wait time, how long Nasdaq queued an order before releasing it for execution. Nasdaq said that would allow "market participants that are less concerned with time to execution to receive executions at the midpoint of the NBBO, while deemphasizing speed as a factor in achieving the execution."[14] And by "limiting interaction of Midpoint Extended Life Orders to other Midpoint Extended Life Orders," Nasdaq hoped M-ELO would mitigate "the impact that these orders will have on the market."[15]

At first Nasdaq held all M-ELO orders for a half-second before releasing them to trade with any other M-ELOs in the book. In 2020 Nasdaq shortened the "holding period" time-in-force term from a half-second for all stocks to 10 milliseconds, reducing the holding period by 98%.[16] Nasdaq proposed in December 2022 changing M-ELO again to use a variable time-in-force term for each stock. The new holding period would vary in a range from .25 to 2.5 milliseconds. Nasdaq planned to use an AI deep learning model to set the exact value every 30 seconds stock-by-stock during the trade day. Nasdaq called its latest version of the order type "Dynamic M-ELO." After a long controversy the SEC approved Dynamic M-ELO in September 2023. In late January 2024 Nasdaq announced it would start rolling it out on February 26. Nasdaq later postponed Dynamic M-ELO to begin on March 25.

The specific AI deep learning method Nasdaq runs to set Dynamic M-ELO's variable time-in-force is reinforcement machine learning. Reinforcement learning can be thought of as a series of trials and rewards that direct a system toward a favored outcome. At first the technology works through random trials as it processes data of some kind. Positive or negative feedback on those random trials directs it in subsequent trials to improve its results and approach that favored outcome. Developers often prepare data they use specifically for these trials. The trial data is called "training data." After the system is calibrated - "trained" - through trials and rewards on the training data, it's ready for production.

Dynamic M-ELO's favored outcome was to find a model that would, depending on market conditions, vary holding periods between .25 and 2.5 milliseconds and improve *aggregate* participant order fill rates and trade markouts beyond the fixed 10 millisecond version of the order type.

---

[13] SEC Release 34-81311, page 13.
[14] SEC Release 34-81311, page 5.
[15] SEC Release 34-81311, page 6.
[16] SEC Release 34-88320.

A "fill rate" is just the portion of units available to trade that do trade. On a stock market, if an order has 1,000 shares available to trade and 500 of them do trade, the fill rate is 50%. A "markout" compares a trade price to the price of the instrument at some point in the future. If prices move your way, you have a good markout. If prices move against you, that's a bad markout. And if prices don't move at all, you're flat. An important point is that a markout is zero-sum. When a buyer on a trade has a good markout the seller has a bad markout and vice versa. However much prices move my way is the exact amount prices move against you. There's always a winner and a loser unless prices stay flat. Nasdaq directed its model to improve fill rates and markouts collectively, averaged across all participants.

To prepare a research prototype, Nasdaq gathered its own historical M-ELO order, trade, and market data from the first quarter of 2022. This historical order and trade data necessarily reflected the 10 millisecond M-ELO then in operation. To train the model, Nasdaq changed the data to "show what the trading environment *would have been* like at timers other than 10ms."[17] Nasdaq also chopped the data down for its test. It's unclear exactly how Nasdaq did that. In one place Nasdaq said it limited its training data to "380 symbols that represent a subset of the 6257 [sic] that are actively traded with M-ELO," a group that included "67% of the current M-ELO volume."[18] In another place Nasdaq said that 67% and those 380 stocks "reflect both actively-traded and thinly-traded securities."[19] So who knows? At least to me, it was also unclear what data Nasdaq would use to train the system in production. The day before the SEC published its approval order Nasdaq said it would use "the same formulation it applied to the initial training set" and use symbols "that collectively account for approximately 67 percent of M-ELO shares traded during the period covered by the retraining" in its production implementation, but no word on exactly how Nasdaq would choose them.[20]

Of course, the world changes through time. Reinforcement machine learning models can try and adapt to those changes if developers "retrain" parts or all of a model with updated data. In its research prototype, and to evaluate its model day-by-day and week-by-week through its 2022Q1 training data, Nasdaq said it had "two retraining independent schedules: one for daily retraining and another one for weekly retraining"[21] and reported an "average combined volume weighted improvement of 31.1%"[22] over the fixed 10ms M-ELO version of the order type. After repeated prodding and some confusion over its proposed production training schedule, Nasdaq clarified it would retrain parts of the model once a week in production but not daily.

The point is these models depend on how developers select, generate, sample, characterize, and update data through time to feed them and how often the models are trained. Training data is curated and shaped. Models are adapted and configured to it.

With deep learning, market centers have to decide:

- What the deep learning method will do. Will it decide a neutral function like a circuit breaker value, something which (likely) affects all participants equally? Will it decide a presumptively neutral function, but which in some circumstances could affect participants unequally? Are there zero-sum decisions, as with trade-by-trade results, where there are winners and losers?;
- The factors which direct the deep learning method. What feedback directs the model to a result? Is it a specific or collective goal and how is it measured?;
- Training data the deep learning method will use, both in development and in production, and how that data is curated and maintained;
- How frequently the deep learning model retrains, and the extent of its retraining.

---

[17] SEC Release 34-98321-ex3a, Diana Kafkes et al., Nasdaq, Inc., "Applying Artificial Intelligence & Reinforcement Learning Methods Towards Improving Execution Outcomes," October 10, 2022 ("White Paper"), page 12 (Exhibit 3A, page 69). Emphasis in original. Available at https://www.sec.gov/files/rules/nasdaq/2023/34-98321-ex3a.pdf.
[18] White Paper, pages 14-15 (Exhibit 3A, pages 71-72).
[19] SEC Release 34-98321, page 8.
[20] Letter from Nasdaq, Inc. dated September 6, 2023, page 2. Nasdaq letters are available at https://www.sec.gov/comments/sr-nasdaq-2022-079/srnasdaq2022079.htm.
[21] White Paper, page 17 (Exhibit 3A, page 74).
[22] White Paper, page 22 (Exhibit 3A, page 79).

Data elements

In response to commenters on the Dynamic M-ELO rule filing,[23] Nasdaq supplied a list of 142 data elements Nasdaq said the model used "that capture market dynamic information about the continuous book and recent M-ELO activity, as well as other information associated with how the timer impacts the simulation [training] environment."[24] Though Nasdaq provided several versions of the list during the rule filing process - versions which featured undefined idiosyncratic terms, changing descriptions, vanishing cells, apparent inconsistencies and redundancies, and mysterious and varying color schemes[25] - the list was clear enough about one thing from the start. Nasdaq kept track of what investors did in each stock from as much as five days ago, and used that history to influence its trade day decisions. It also became clear that in some circumstances Nasdaq might use investor behavior from weeks or months ago to affect the model's decisions.[26] This is important. To my knowledge no U.S. stock exchange has ever used past participant behavior - from seconds ago, minutes ago, days ago, or even weeks or months ago - to decide a new order's terms.

With deep learning, market centers should:

- Fully disclose data elements the deep learning model trains with (during or outside market hours), and disclose all data elements it uses to make determinations during trading hours;
- Define data elements using standard industry terminology, and fully explain any coinages or idiosyncratic terms.

Statistics

During a trade day and in retraining, Nasdaq keeps track of behavior by computing various order and trade statistics like sums, averages, and percentiles by stock. Nasdaq also computes measures of distribution spread and shape in this data like standard deviation, skew, and kurtosis, also by stock. Most of the 142 data elements are those kinds of statistics and Nasdaq calculates them over time periods ranging from the previous 30 seconds to the last five days. As examples, they include figures like: "the proportion of BUY orders compared to all orders in the M-ELO orderbook over the last day, on a per ticker basis, during regular trading hours"; "the count of unique firms that placed BUY orders in the M-ELO orderbook over the last day, on a per ticker basis, during regular trading hours"; "Maximum difference between the best ask and best bid prices over the last trading day, per ticker, including after-hours trading"; "The total buyside trade quantity that came from the simulated trades made by the agent's action over the last 90 seconds, per ticker"; and "The standard deviation of buyside trade quantity that came from the simulated trades made by the agent's action over the last 5 minutes, per ticker."[27]

Nasdaq said that "The System will use only aggregate statistics across all participating parties, and it will not make changes that are individualized to a specific participant or investor,"[28] but didn't note whether and how Nasdaq contends with sparse data or grossly uneven data from which it calculates those statistics. Because of sparse or cockeyed data, I believe Nasdaq can in fact make decisions specific to a participant or investor, even to their disadvantage.

Anyone familiar with markets and market data can see the problem. Activity varies tremendously across firms and traded instruments. In the stock market, Apple or Google will see hundreds of firms and millions of orders and trades during the day. In the futures markets, a front month can be many times busier than other months. In all markets, as you move to downlist instruments the number of participants shrinks and activity slows to a trickle. Among the 6,257 stocks Dynamic M-ELO traded, thousands of downlist names might have a small handful - even just one or two - of participants, and just a sprinkling of

---

[23] Leuchtkafer 1, page 1 and Themis, page 2.
[24] White Paper, page 27 (Exhibit 3A, page 84).
[25] See Leuchtkafer 2, pages 8-9, Leuchtkafer 3, pages 8-9, Leuchtkafer 4, pages 8-9, and Leuchtkafer 5, pages 1-2.
[26] See Leuchtkafer 4, page 11.
[27] SEC Release 34-98321-ex3b, available at https://www.sec.gov/files/rules/nasdaq/2023/34-98321-ex3b.pdf.
[28] Letter from Nasdaq, Inc. dated May 18, 2023, page 2.

orders and trades Dynamic M-ELO can use to calculate its statistics. As for participants, over a day a busy professional like a market maker might submit millions of orders while even an active investor might submit just a few dozen.

Let's explore this with an admittedly simple analogy but which shows the point. Suppose someone calculates the average weight of everyone who happens to be in a first grade classroom at a particular time. They might get a reasonable sense of how the classroom measures up if there are 25 six year-old kids in the room and no one else. The average will probably be a pretty reasonable description of most of the kids. One reason is that there are a fair number of kids in the room. Another reason is that for the most part first grader weights usually range in a certain way. If on the other hand they calculate the average of a first grade classroom with just four students in the room, the average is a less reliable description. Two small kids could weigh 35 pounds each and two larger kids could weigh 85 pounds each. The average is 60 pounds but they won't find anyone in the room anywhere near that weight. They also won't have much insight about first graders overall, if that's what they're looking for. If there's just one student in the room, they won't know anything but what that one student measured. And if New York Jets offensive lineman Mekhi Becton happened to drop by for show-and-tell when any of these statistics were calculated, at 363 pounds the statistics on the room will veer wildly toward Becton and away from the six year-olds.

The classroom in the analogy is a stand-in for a stock. The people in the room are stand-ins for the participants in a stock over any particular time period. Weight statistics are like the statistics on participant orders and trades. If all participants have generally similar order and trade behavior when Nasdaq computes its statistics on a stock, the statistics will describe something meaningful, something they have in common. If the participants and their behaviors are very different from one another, or if even a single participant and its behavior are very different from the others, the statistics might not say much that's meaningful - the statistics might even tilt wildly toward one and away from the others. All this can be aggravated because many of Nasdaq's statistics are further divided by side, buy and sell. But as described in Nasdaq's filing, Dynamic M-ELO will eat all this statistical data, whatever it is and however sparse or lopsided the underlying orders and trades are, when it decides how to handle participant orders.

And so I believe Nasdaq's way of keeping track of what investors have done over time could very well influence Dynamic M-ELO to make order changes "individualized to a specific participant or investor" when it uses these statistics to affect fill rates and markouts. Because markouts are zero-sum, those changes could leave a specific participant worse off and its counterparty better off on a trade than if Nasdaq simply stayed out of it. So far as I saw, Nasdaq never addressed this. The SEC ignored it too. I believe this is also entirely new ground in U.S. stock markets. The SEC now allows exchanges to pick and choose winners and losers with a technology I believe could easily favor some participants or participant types over others.

If market centers use participant data in a deep learning method, they should:

- Manage widely variable behaviors among participant types and instruments, and disclose what they do;
- Manage sparse or cockeyed data and whether and how they might cause their deep learning method to make decisions that are "individualized to a specific participant or investor," and disclose what they do.

Sampling

Nasdaq said its way of chopping down M-ELO's 2022Q1 order and trade data to train Dynamic M-ELO on 380 stocks instead of all 6,257 stocks was consistent with the accepted practice of population sampling.[29] I think Nasdaq plainly didn't understand the math, or didn't want to. Statisticians sample populations, and they do it in areas like voter and opinion polling and manufacturing. They do it when it's difficult or even impossible to survey an entire population, and they go to extraordinary lengths to design their sampling methods and verify these methods draw data reflecting the populations from which they are drawn. It's an election year and voter polling is a good example. For most elections it's impossible to survey every voter, so statisticians design sampling methods they believe reflect the voting population-at-large. And then prospective survey respondents are identified and screened to make sure those respondents fit the sampling method itself, that is, that the sample data collected reflects - as best it can - the population-at-large in all material aspects.[30]

In its filings, Nasdaq didn't describe doing any of this or explain its reasons for sampling in the first place. Instead, Nasdaq simply said it used 380 stocks (excluding 94% of the stock population-at-large) when it developed the system, and will use the same or a similar group of stocks to retrain the system in production. It didn't say whether and how the 380 stocks it selected are consistent in kind and proportion with the total population of 6,257 stocks traded with M-ELO orders in 2022Q1. From what Nasdaq described in at least one place, the system was developed, tested, evaluated, and will run in production using a hand-picked group of active stocks to train,[31] and not something a statistician would recognize as a truly representative sample of participants and stocks. I asked about that.[32] Nasdaq didn't respond. Nasdaq also said it summarily excluded all low activity firms from its review of Dynamic M-ELO's performance "to avoid their data distorting the results."[33] It didn't say whether the low activity firms are participant types consistent in kind and proportion with the total population of participants, and Nasdaq hasn't shown what Dynamic M-ELO's performance is when those participants are included in its review.

Sampling problems are a known concern with AI. The National Institute of Standards and Technology ("NIST") warns "Computational and statistical biases can be present in AI datasets and algorithmic processes, and often stem from systematic errors due to non-representative samples."[34] Experts have also weighed in specifically on sampling and machine learning. In a 2021 paper researchers at Stanford wrote "A data set sampled from a certain population is biased if the subgroups of the population are sampled at proportions that are significantly different from their underlying proportions"[35] and "Applying machine learning algorithms naively to biased training data can raise serious concerns and lead to controversial results." Further, "a model trained from biased data tends to favor oversampled subgroups by achieving high accuracy there while sacrificing the performance on undersampled subgroups." Those researchers then described methods to compensate for these factors, none of which Nasdaq said it used.

Another concern is that the training data will be appropriate through time. In its filings, Nasdaq didn't describe how it will maintain its sample of stocks in production and what criteria it will use to add and delete stocks other than that it would "test all activity for those symbols that collectively account for approximately 67 percent of M-ELO shares traded during the period covered by the retraining."[36] Since Nasdaq will retrain weekly, does that mean it will update the sample once a week? What will Nasdaq do when the next meme stock frenzy hits? What will Nasdaq do when another Facebook IPOs? What will Nasdaq do when companies on its list merge, get acquired, go bankrupt, or when they have dramatic

[29] Letter from Nasdaq, Inc. dated September 6, 2023, page 4.
[30] Even so, sampling is not perfect. It's why voter surveys note their margin of error, an estimate of how chance error might affect results.
[31] 380 of 6,257 stocks are very unlikely to include 67% of traded volume unless they have been hand-picked (as opposed to picked randomly).
[32] Leuchtkafer 5, pages 4-5.
[33] SEC Release 34-98321, note 40, page 21.
[34] National Institute of Standards and Technology, U.S. Department of Commerce, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)", January 2023, page 18.
[35] Jing An, Lexing Ying, and Yuhua Zhu, "Why Resampling Outperforms Reweighting for Correcting Sampling Bias with Stochastic Gradients" (2021). Available at http://web.stanford.edu/~lexing/resw.pdf.
[36] Letter from Nasdaq, Inc. dated September 6, 2023, page 2.

adjustments from splits, reverse-splits, spin-offs, or other corporate actions? And will Nasdaq clean up dirty data from algorithms on the fritz, fat fingers, busted trades, or system bugs before including it in its training data? All these are concerns with a market data training set in any case, but the smaller the subsets of stocks and participants in the training set the more acute these questions become. I asked about that too.[37] Nasdaq didn't respond.

Most important, statisticians sample populations when it is unworkable or impossible to collect data from an entire population. *But market centers have the entire populations at their fingertips*. The populations here are the participants and the instruments in which they trade and all relevant market data. Without good cause I believe there isn't a reason to "sample" this data at all, and there are important reasons not to.

In training a deep learning method, market centers should:

- Use all the data to train. They have it and they should use it. If they don't use all the data to train, they should explain in detail why they don't. They should then show results on all data alongside results from their proposed sample so regulators and participants know how the model behaves and the effect of the market center's decision not to use all data;
- Show all their results. Regulators should not let market centers cherry-pick to avoid "data distorting the results." On the one hand Nasdaq said Dynamic M-ELO will "operate pursuant to a mathematical algorithm from which it cannot deviate"[38] and its results on "day 1 should be the same as they would be on day 2 [between retraining],"[39] and on the other that it excluded what it called "statistical noise"[40] or "outlier data"[41] from its reviews. If the system operates with a mathematical algorithm from which it cannot deviate, that is utter malarkey. Any "statistical noise" or "outlier data" is an integral consequence of how it works and what can happen.

Punted

Stock exchanges have a statutory requirement to ensure their rules are "not designed to permit unfair discrimination between customers, issuers, brokers, or dealers."[42] Contract markets have a statutory responsibility to "promote fair and equitable trading."[43] With current market technology, predictable and invariant trading systems always treat orders the same way according to market rules, and it is straightforward to ensure the technology doesn't unfairly discriminate and trading is fair and equitable. It is also straightforward enough for market centers, participants, and regulators to police how these systems work and that they work according to the rules and expectations. The systems are transparent and deterministic. A concern with AI is not just that a market center somewhere might covertly use it to favor one participant type over another. It's that the technology could do it anyway. That's what the NIST warned about with "Computational and statistical biases" and Stanford researchers warned about with "biased training data [that] can raise serious concerns and lead to controversial results."

In its early filings for Dynamic M-ELO, Nasdaq reported that it ran an analysis and found no "systematic-biased execution towards any *one firm*."[44] That requirement is obviously implied by the Exchange Act, but the statute is broader than that and requires no unfair discrimination between participant types. If Nasdaq used 380 "actively traded" stocks to train and evaluate Dynamic M-ELO, I believe excluding 5,877 less actively traded stocks (6,257 - 380 = 5,877) or excluding any participants at all from a bias analysis is no way to confirm the system is free from unfair discrimination. In response to comments, Nasdaq later reported that it tested for bias among participant types but didn't say what

[37] Leuchtkafer 5, page 5.
[38] SEC Release 34-98321, page 23.
[39] SEC Release 34-98321, page 16.
[40] Letter from Nasdaq, Inc. dated September 6, 2023, page 5.
[41] Letter from Nasdaq, Inc. dated Nasdaq September 6, 2023, page 5.
[42] Section 6(b)(5) of the Securities Exchange Act.
[43] Section 5(d)(12)(B) of the Commodity Exchange Act.
[44] White Paper, page 26 (Exhibit 3A, page 83). Emphasis added.

factors it used to classify them or what classifications it used, and it didn't show the numbers. Nasdaq did say that it detected no "material variations" or "unreasonably disproportionate" benefits among participant types,[45] but I believe these qualified and unsubstantiated conclusions raise more questions than they answer. The SEC and the public should have seen all the data from these tests.

But the SEC punted. It's punted on rule filings in the past and got sued. The courts were savage about it. In *Susquehanna Int'l Grp., LLP v. SEC* (D.C. Cir. 2017), in overturning a SEC approval order the D.C. Circuit wrote "the SEC should have critically reviewed [the SRO's] analysis or performed its own," and quoted from *NetCoalition v. SEC* (D.C. Cir. 2010) where it found a "lack of support in the record" and that "The SEC had tried to rely on statements by the self-regulatory organization, but we saw 'little' supporting value in the 'self-serving views of the regulated entit[y].'" Later the SEC itself wrote "The D.C. Circuit's *Susquehanna* Opinion makes clear that relying on such [SRO] representations, without more, is insufficient. Rather, the Commission must critically evaluate the representations made and the conclusions drawn by [the SRO]."[46] Instead of critically evaluating Nasdaq's analyses and representations the SEC's Dynamic M-ELO approval order says "Nasdaq represented" and "Nasdaq concluded" and "Nasdaq explains" and shows no signs - none - the agency itself examined the data or performed its own analysis.[47] Maybe it doesn't know how.

This technology could be one of the most notable developments in our markets in decades. In light of *Susequehanna* however it seems as if the SEC shrugged and dared someone to sue it over Dynamic M-ELO. I'm a D-list personality in these matters[48] and it won't be me, but that's the current pose of the country's "investor advocate."

For a deep learning method:

- Market centers should do comprehensive bias studies and disclose detailed results, broken out into important categories like participant type, instrument type, and participant and instrument activity tiers. Regular bias testing - at minimum whenever a deep learning model is retrained - should be required and disclosed;
- Regulators should critically evaluate bias studies for any sign of unfairness and not rely on a market center's "self-serving views."

Established or pre-determined, and non-discretionary

The late 1990s were a reform era for the stock markets. Responding to the rise of automated and semi-automated markets and to an historic quote-rigging scandal at Nasdaq (which was then owned and operated by FINRA's predecessor), the SEC passed several rules that led to what the stock market looks like today. Among those rules, in 1998 the SEC approved a new regulation called Regulation ATS ("Reg ATS"). Within Reg ATS, Rule 3b-16 further refined the definition of a "stock exchange" under the Exchange Act. It says a stock exchange "(1) Brings together the orders for securities of multiple buyers and sellers; and (2) Uses established, non-discretionary methods (whether by providing a trading facility or by setting rules) under which such orders interact with each other, and the buyers and sellers entering such orders agree to the terms of a trade."

For me and Dynamic M-ELO, the matter hinged on whether a stock exchange could implement something like Dynamic M-ELO. Was it an "established, non-discretionary method," both as to the original intent of the phrase and with its plain meaning today? I believe the phrase requires that exchanges use transparent and deterministic rules for how an exchange handles and matches orders. And with reinforcement machine learning, were *buyers* and *sellers* agreeing to the terms of a trade or was the exchange itself determining terms, the way a broker might? The matter also hinged on how the SEC had

[45] SEC Release 34-98321, page 21. See also pages 38-39.
[46] SEC Release 34-85121, page 2.
[47] SEC Release 34-98321, page 39.
[48] "Concerned" - *Barron's*; "Feisty" - *Bloomberg*; "Regular" - *Financial Times*; "Smarmy, know nothing know it all" - *Hudson River Trading*.

previously interpreted what "established, non-discretionary methods" meant. ("Established, non-discretionary methods" is very similar to "pre-determined non-discretionary automated trade matching and execution algorithm," which is how the Commodity Exchange Act defines "trading facility."[49])

The Exchange Act defines an exchange as "a market place or facilities for bringing together purchasers and sellers of securities or for otherwise performing with respect to securities the functions commonly performed by a stock exchange as that term is generally understood."[50] As was generally understood at the time of enactment, there's no room for a "stock exchange" to routinely intervene and set the material terms of buy and sell orders according to the exchange's own private determinations. Stock exchanges have the *regulatory* authority to intervene, such as to bust or adjust trades in extraordinary circumstances, but aren't allowed the *business* control over participants to change their order terms based on the exchange's own varying and unpredictable judgments.[51] There's no suggestion anywhere an exchange can be anything more than a strictly neutral forum for buyers and sellers to conduct their business together, and according to their own terms.

That neutrality has been underscored time and again by the SEC in think pieces, policy statements, and new rules. In a 1994 report called "Market 2000" the SEC wrote "The Commission has determined this function [an exchange] to be the provision of a trading market that is designed, whether through trading rules, operational procedures, or business incentives, to centralize trading" which, again, defines an exchange as a neutral facility tasked only with "the provision of a trading market" for buyers and sellers to conduct their business according to their own terms.[52] In that context, four years after Market 2000 the SEC adopted Reg ATS and Rule 3b-16, and again there is no suggestion that an exchange is anything but a neutral marketplace that "brings together the orders" of its participants and uses "established, non-discretionary methods" under which those buyers and sellers will conduct their business according to their own agreed terms. When Reg ATS was adopted, there wasn't even a hint an exchange would routinely alter the material terms of a participant order based on its own unpredictable judgments and its own undisclosed data, as Nasdaq proposed with Dynamic M-ELO.

The record continues and expands. When the SEC approved a new exchange called IEX in 2016, the SEC very directly took on the question of what "established, non-discretionary methods" means. There was a lot of controversy at the time whether one of IEX's proposed order types was an impermissible discretionary method for an exchange. In its most exhaustive analysis of Rule 3b-16 in modern memory, the SEC wrote "IEX has thus encoded in its rule the *totality* of the discretionary feature of its proposed discretionary peg.... the Commission does not believe that the *hardcoded* conditionality of the IEX proposed 'discretionary' peg order type provides IEX with actual discretion or the *ability to exercise individualized judgment* when executing an order."[53] Further, because the controversial feature behind the order type was "based on a *pre-determined*, *objective* set of conditions that are detailed in IEX's proposed rule" the SEC would allow it.[54]

Merriam-Webster's defines "totality" as "the quality or state of being total" and "total" as "comprising or constituting a whole." For all time in the computer age "hardcoded" has meant that programmers specifically tell a system what to invariantly do in a circumstance. Merriam-Webster's defines "individualized" as "to make individual in character" and "judgment" as "the process of forming an opinion or evaluation by discerning and comparing." It defines "pre-determined" as "to determine beforehand" and "objective" as "expressing or dealing with facts or conditions as perceived without distortion by personal feelings, prejudices, or interpretations."

Because the results of reinforcement machine learning methods are unpredictable, uncertain, and so complex they are unexplainable and unauditable, they can't meet the plain language used by the SEC in

---

[49] Commodity Exchange Act, Section 1(a)(51)(A)(ii).
[50] Securities Exchange Act, Section 3(a)(1).
[51] What about pegged orders? Pegs are deterministic. Exchanges predictably and mechanically apply them. An exchange can't peg orders whenever it wants to, especially if it can't say where and why it will do it.
[52] SEC, "Market 2000, An Examination of Current Equity Market Developments," Study III - 12, (1994).
[53] SEC Release 34-78101, pages 44-45. Emphasis added.
[54] SEC Release 34-78101, page 41. Emphasis added.

its IEX approval. Machine learning results aren't hardcoded or pre-determined. They are individualized when inputs veer one way or another because of a participant or participant type, and they deal with facts and conditions at the moment through prejudices and interpretations they derived during training - training data that itself depends entirely on judgments the exchange made when it selected, generated, sampled, characterized, and updated that data.

Later, IEX proposed a new order type, and the SEC again visited the question of what "established, non-discretionary methods" means. In its approval of that order type the SEC doubled-down on what it had said before, emphasizing its previous ruling. It also clarified the principles of transparency and determinism with another standard. It wrote that the feature distinguishing the order type and a source of controversy, was allowable because it was based on a "'pre-determined, objective set of conditions that are detailed in IEX's [rules]' and which *any market participant can thus recreate on its own*."[55]

In contrast to IEX, as approved by the SEC, Nasdaq's rulebook doesn't set out the "totality of the discretionary feature." I believe it can't, because the totality changes week-to-week and even minute-to-minute. It doesn't define the "hardcoded conditionality" of its feature - again, I believe it can't. A market participant won't be able to "recreate on its own" what Dynamic M-ELO does - participants can't and it's not clear anyone can, ever. I believe Dynamic M-ELO will also "exercise individualized judgment" such that it can set a different time-in-force for the very same order presented in the very same market conditions on, say, August 21 than it set on May 15, depending on the system's undisclosed judgments of market conditions and participant behavior from even days or weeks in the past, characteristics of this technology Nasdaq has not disputed.

For deep learning:

- Regulators should propose new regulations or new interpretive guidance for this technology;
- Regulators shouldn't pretend that regulatory frameworks enacted decades ago make room for AI. As I discuss below, the SEC took a cynical shortcut to accommodate a registrant and approve a rule filing. There's little enough faith in government. Do the right thing.

Frisbee

Precedent is for suckers. The SEC made that clear in its Dynamic M-ELO approval order. The agency's prior reasoning for another exchange didn't matter.[56] Facts and circumstances are different, the SEC said. The principles are apparently different too, so the SEC laid down new principles. It agreed that Dynamic M-ELO was "so complex that its complete details are, for most intents and purposes, not readily intelligible."[57] It also agreed that "it would be immensely difficult for the Exchange or any market participant to precisely predict"[58] what Dynamic M-ELO would do. Nevertheless the SEC would allow "an unavoidable degree of uncertainty with respect to the use of these order types."[59]

So I believe exchanges could save a lot of money - AI programmers aren't cheap, and the infrastructure to run AI isn't cheap either - if they just used monkeys. An exchange could set up targets for different order terms and let monkeys throw darts at them to set values for those terms. Under the SEC's old exchange standards that would be a problem. Is it a problem under the SEC's new standards? Let's look at a table.

[55] SEC Release 34-89686, page 28. Emphasis added.
[56] SEC Release 34-98321, note 60, page 29.
[57] SEC Release 34-98321, page 30.
[58] SEC Release 34-98321, page 30.
[59] SEC Release 34-98321, page 30.

SEC Exchange Standards

Old Standard	Can You Use Monkeys?	New Standard	Can You Use Monkeys?
"encoded in its rule the totality of the discretionary feature"	🚫	"so complex that its complete details are, for most intents and purposes, not readily intelligible"	🐒
"based on a pre-determined, objective set of conditions that are detailed in [the] proposed rule"	🚫	"an unavoidable degree of uncertainty with respect to the use of these order types"	🐒
"'pre-determined, objective set of conditions that are detailed in [the rules]' and which any market participant can thus recreate on its own"	🚫	"it would be immensely difficult for the Exchange or any market participant to precisely predict [what it will do]"	🐒

Can a monkey do a good job? Nasdaq said a random selection of holding periods between .25 and 2.5 milliseconds showed better results than several different fixed holding periods it tested, so there's good reason to believe that, yes, in fact, a monkey can do a good job.[60]

Deep learning methods can be as transparent, deterministic, auditable, and accountable as a monkey. If anyone believes, as I do, that automated markets were a huge advance over manual markets because automation made markets transparent, predictable, and auditable in ways manual markets never were, deep learning undercuts that progress as much as a monkey would.

Audit trails and disclosures

If regulators approve deep learning in market center trading systems, they should update regulations to account for it. Deep learning is as large a shift as automated markets were in the first place. A regulator shouldn't pretend that regulations adapted to one technology and its distinctions and expectations can be twisted to technology that reverses those distinctions and expectations, however much anyone believes in the change.

As an example, regulators need to define appropriate audit trails for this technology. Market center audit trails today show data exactly as it arrived and record material events as the data is processed. But today's processing is hardcoded and deterministic. Presumably audit trails will show what a deep learning method determines - its end result - but how will regulators police that it was done correctly? What do regulators need to police these models, in their design, training cycles, and deployment? What do regulators need to reconstruct deep learning model state changes to try and understand why a model did what it did? At minimum market centers should archive every instance of their training data, every instance of their deep learning model states and parameters, and every piece of data they consume or create during the trade day.

Regulators should also insist on appropriate disclosures. These disclosures are important because deep learning methods are so fundamentally different from the market technology of the last half-century.

[60] White Paper, page 25 (Exhibit 3A, page 82), see "Combined Improvement (%)" figures in "Benchmarking AI Model Against Static Timer and Random Timer" chart. Using the "Random Timer" as a proxy for a monkey, the monkey could deliver roughly a 10% improvement over the legacy fixed 10ms holding period and perform better than several other fixed holding periods Nasdaq tested. You can play frisbee with it too. With a roughly 30% improvement over the fixed 10ms holding period the AI model did better than a monkey would, but I imagine is much more expensive to operate and maintain.

Participants need to understand what it means. I believe Nasdaq's descriptions of the technology in its filings and public statements were sprinkled with happy talk, as if an AI fairy was coming to bring candy and toys to everyone who used Dynamic M-ELO. But because Nasdaq directed Dynamic M-ELO to improve a collective outcome - overall average fill rate and markout results - and because markouts are zero-sum on a trade, I believe a participant on a particular trade could come out worse and its counterparty better than if Nasdaq hadn't inserted itself. And because Dynamic M-ELO uses participant behavior in its analyses to try and improve collective outcomes, I believe it's possible a participant's own past behavior could be used to its disadvantage.

In its approval order, the SEC was hear-no-evil about whether a participant's past behavior could be used against it. And so participants might naively believe these methods produce Lake Wobegon results, where every trade is above average. But they don't. And with the staggering complexity of these methods, and how they reconfigure themselves whenever they retrain, market centers should be clear to participants and regulators that specific outcomes are unpredictable, unauditable, and unexplainable, in the moment and through time.

I believe these points should have been disclosed about Dynamic M-ELO:

> *That an exchange will use hidden aspects of its current and past order flow to freely and unpredictably change an order's material terms.*
>
> *That an exchange doesn't have to hardcode rules and procedures for its automated order types. Rules and procedures don't have to be hardcoded at all. Even if market conditions are the same, a deep learning method might do something different in the future.*
>
> *That an exchange can use a participant's current and past behavior to freely and unpredictably change an order's material terms, even if it means worse results for that participant relative to a collective, relative to its counterparties, or relative to what would have happened had the exchange stayed out of it.*
>
> *That an exchange can use a participant's current and past counterparty behavior to change material terms on that participant's order, affecting how it handles the order, even if it means worse results for that participant relative to a collective, relative to its counterparty, or relative to what would have happened had the exchange stayed out of it.*
>
> *That an exchange can in fact exercise variable control, judgment, and discretion over participant orders, and because of the complexity of this control, judgment, and discretion specific results are unpredictable, unauditable, and unexplainable. An exchange can't tell participants or regulators exactly what a deep learning model did and will do and why.*

Sincerely,

R. T. Leuchtkafer

Deep Learning in Public Markets

Unlike today's technology:

- Deep learning is not deterministic. Results are unpredictable, unexplainable, and unauditable;
- Deep learning depends on the data it's fed and how developers select, generate, sample, characterize, and update that data;
- Despite intentions, deep learning can discriminate against participants or participant classes;
- Market centers can't fully explain to regulators or participants specific results. Why deep learning did one thing as opposed to any other can be so complex it's unknowable;
- Because specific results are unpredictable, unexplainable, and unauditable, regulators and participants can't police what market centers do and why.

With deep learning, market centers have to decide:

- What the deep learning method will do. Will it decide a neutral function like a circuit breaker value, something which (likely) affects all participants equally? Will it decide a presumptively neutral function, but which in some circumstances could affect participants unequally? Are there zero-sum decisions, as with trade-by-trade results, where there are winners and losers?;
- The factors which direct the deep learning method. What feedback directs the model to a result? Is it a specific or collective goal and how is it measured?;
- Training data the deep learning method will use, both in development and in production, and how that data is curated and maintained;
- How frequently the deep learning model retrains, and the extent of its retraining.

After choosing data to train a deep learning model, market centers should:

- Fully disclose data elements the deep learning model trains with (during or outside market hours), and disclose all data elements it uses to make determinations during trading hours;
- Define data elements using standard industry terminology, and fully explain any coinages or idiosyncratic terms.

If market centers use participant data in a deep learning method, they should:

- Manage widely variable behaviors among participant types and instruments, and disclose what they do;
- Manage sparse or cockeyed data and whether and how they might cause their deep learning method to make decisions that are "individualized to a specific participant or investor," and disclose what they do.

In training a deep learning method, market centers should:

- Use all the data to train. They have it and they should use it. If they don't use all the data to train, they should explain in detail why they don't. They should then show results on all data alongside results from their proposed sample so regulators and participants know how the model behaves and the effect of the market center's decision not to use all data;
- Show all their results. Regulators should not let market centers cherry-pick to avoid "data distorting the results."

Market centers should test their models for any sign of bias:

- Market centers should do comprehensive bias studies and disclose detailed results, broken out into important categories like participant type, instrument type, and participant and instrument

activity tiers. Regular bias testing - at minimum whenever a deep learning model is retrained - should be required and disclosed;
- Regulators should critically evaluate bias studies for any sign of unfairness and not rely on a market center's "self-serving views."

For deep learning:

- Regulators should propose new regulations or new interpretive guidance for this technology;
- Regulators shouldn't pretend that regulatory frameworks enacted decades ago make room for AI. Regulations adapted to one technology and its distinctions and expectations can't be twisted to technology that reverses those distinctions and expectations, however much anyone believes in the change;
- The SEC took a cynical shortcut to accommodate a registrant and approve its rule filing. There's little enough faith in government. Do the right thing.