From: Darren Lee
To: Comments, Public
Cc: John Pepe; Daniel Rapp; Rick Dales; Michael Mcgrath
Subject: Commentary on FINRA Paper, Artificial Intelligence (AI) in the Securities Industry from Proofpoint
Date: Monday, August 31, 2020 5:05:27 PM

To:
Haimera Workie, Senior Director and Head of Office of Financial Innovation
Kavita Jain, Director, Office of Financial Innovation

From:
Proofpoint Digital Risk and Compliance
Compliance Customer Product Council

In Regards to:
Commentary on FINRA Paper, *Artificial Intelligence (AI) in the Securities Industry*

Greetings, FINRA Office of Financial Innovation:

Proofpoint is a leading cybersecurity and compliance technology vendor for the financial services industry with a good number of our customers being FINRA members. As a company focused on innovation and the needs of our customers, we keep up to date on FINRA publications and announcements that may affect product offerings and roadmaps. FINRA's very informative paper regarding *Artificial Intelligence (AI) in the Securities Industry* is of special importance to our customers since many are beginning to use Proofpoint's AI tools for their regulatory compliance programs.

To help guide future direction, Proofpoint periodically convenes with our Compliance Customer Product Council ("CPC") which is comprised of the top broker-dealers in the US and Canada. In collaboration with Proofpoint and our CPC, we are submitting commentary on FINRA's paper regarding *Artificial Intelligence (AI) in the Securities Industry*. Sections and page numbers referenced in the below commentary relate to the PDF version of the FINRA Paper.

Proofpoint and our Compliance Customer Product Council welcome the opportunity to participate in helping draft industry guidance. The below feedback is structured in ten separate sections representing our collective feedback to FINRA's paper.

I. AI Models Learning from Each Other in Surveillance and Monitoring

Relating to how "*AI trading models across the industry may start to learn from each other - potentially leading to collusive activity, herd behavior, or unpredictable results*", under the *Portfolio Management and Trading* section starting on page 7 of the FINRA Paper (PDF version), the Proofpoint CPC may see some benefits in models learning from each other when applied to surveillance and monitoring.

AI models that interact with the market directly are active market participants. As such, if that model is also learning, then it follows that it will learn from all market participants, including other dynamic models. As the frequency of transactions that will result from models increases, then they will probably have a more considerable significance on models and market participants.

But there is also the opportunity for this effect to be deliberately applied across the industry to reduce risk and operational cost. That is in the sphere of surveillance, monitoring, and fraud detection. Models which are learning from content to identify risky communications will benefit if they can learn from each other and larger sets of communication sets. This creates an opportunity to improve the operation of the market by reducing risk by allowing models to learn from each other. It suggests that models can outperform current practice because, where they are not a source of corporate competitive advantage (Porter, 1996), by being able to learn from broader data sources, longitudinally cooperatively. This sort of co-operation also suggests itself as a source of community archiving, which will allow a considerable opportunity to reduce cross-industry cost. Such a solution will require an industry-wide solution that can perform at scale and with the type of access and security offer the industry the requisite privacy, security and access control. These technical challenges have, of course, all been resolved. Community use of utility archiving and compliance will allow us to produce superior models at a lower cost to the industry.

Porter, M. E. (1996) ‘Competitive Advantage, Agglomeration Economies, and Regional Policy’, *International Regional Science Review*. SAGE Publications Sage CA: Los Angeles, CA, 19(2), pp. 85–90. doi: 10.1177/016001769601900208.

II. Using Surveillance and Monitoring Systems to Remove False Positives

Relating to *Surveillance and Monitoring* under the *Operational Functions, Compliance and Risk Management* section starting on page 8 of the FINRA Paper (PDF version), the Proofpoint CPC agrees that AI tools could "significantly reduce the number of false positives, which in turn, free up compliance and supervisory staff time to conduct more thorough reviews of the remaining alerts, resulting in higher escalation rates." However, much of the discussion of AI has focused on improved detection of ‘violations’. Equally, if not more valuable, is the identification of non-relevant alerts/content, or ‘false positives’. AI can be effectively employed in these situations to automatically eliminate obvious false positives, leaving the more difficult to judge cases for expert personnel. The reduction of volume requiring review will allow member firms to focus efforts on the more critical aspects of compliance.

To be effective and relevant, AI models require a lot of data and examples. According to our CPC members, 99.5% of surveillance matches are 'non-violative' which leaves only about 0.50% of the population to train the many 'inclusion' models that attempt to find the 'needle in the haystack'. Therefore, using AI to first eliminate obvious false positives ('exclusions') may be a prudent approach to ensure that potential violations are not missed, which then allows human reviewers to analyze the 'cleaned' 0.05% population leftover.

III. Model Explainability and Statistically Sufficient Sampling

Relating to *Model Explainability*, under the *Model Risk Management* section starting on page 11 of the FINRA Paper (PDF version), the Proofpoint CPC agrees that many AI and machine learning (“ML”) models, and in particular 'deep learning' models, behave in a 'black box' fashion. While the predictive performance of the model provides accurate and useful results, the reason that any given prediction is made isn't clear. Ongoing research in the area of model explainability is showing some promise, but practical application in most cases will be in the intermediate to long term.

In those areas where ML is applied in non-real-time applications, another aspect should be considered: *Does the model perform at or above human expert level performance?* The application of ML to surveillance and monitoring is a prime example of this. To ascertain whether a model is performing at or above human level performance, the technology application should route statistically sufficient samples of potential violations to both compliance/review personnel and the machine learning system. Results of both reviews can be compared and validate/certify ML model performance.

Member guidance on 1) specifications as to what constitutes statistically sufficient samples, and 2) a model validation report format that auditors should expect, would be helpful to the Proofpoint Compliance Customer Product Council.

IV. Data Governance: Applicability of Data Sources

Relating to *Data Governance*, starting on page 13 of the FINRA Paper (PDF version), the Proofpoint CPC adds the following commentary on the Applicability of Data Sources:

Data sources used for training must align with the data sources that they will be applied to, as significant variances can exist. For example, discussions on social media and chat systems rely heavily on short-forms, acronyms, and emojis. Applying models trained on longer-form communications to these sources will be inaccurate. Similarly, textual context extracted from images or audio have biases introduced by the OCR or speech to text solutions (which resolve ambiguity based upon the most common result for the context). These variances in the source data should be considered when deciding if a ‘unified’ model versus a ‘source-specific’ model is appropriate.

V. Data Governance: Applicability of Data Sources

Relating to Data Security under the *Data Governance* section starting on page 13 of the FINRA Paper (PDF version), the Proofpoint CPC adds the following commentary:

While security techniques, such as encryption, are commonplace for longer term data storage, most analytic techniques require access to plain-text content. Depending upon the use case, the scope of analysis will often require a subset of training data to be stored without encryption. Specialized controls are needed for these analysis environments, given the greater risks involved.

VI. Data Governance: Cybersecurity Considerations

Relating to *Cybersecurity Considerations* under the *Data Governance* section starting on page 13 of the FINRA Paper (PDF version), the Proofpoint CPC adds the following commentary:

In addition to bias that comes from the sourcing and labelling of data, it is reasonable to assume that, with the growing reliance on AI, cybercriminals will seek to influence models by manipulating the data to intentionally introduce bias. In addition to access controls and security procedures, generating fingerprints that represent the data at different stages and storing them independently can be critical to detect unintended changes.

VII. Customer Privacy

Relating to the *Customer Privacy* section starting on page 15 of the FINRA Paper (PDF version), the Proofpoint CPC adds the following commentary:

AI techniques are frequently used to build a rich profile on an individual. For this to be useful, it needs to ultimately tie back to the person – even if standard PII (such as a social security number) is not used as the identifier. The richer the data sources, the easier it becomes to make inferences about personal characteristics. For example, access to banking and credit transactions may be sufficient to infer gender, race, country of origin, sexual orientation, and education. If these attributes are identified and stored, even though they may not critical to the business problem being addressed, they create risk of abuse.

VIII. Cybersecurity: Additional Considerations

Relating to *Cybersecurity* under the *Additional Considerations* section starting on page 18 of the FINRA Paper (PDF version), the Proofpoint CPC agrees that firms would benefit from incorporating cybersecurity as a critical component of the evaluation, development, and testing process of AI-based application and would like to add that, given the rapidly changing and targeted nature of modern cybersecurity attacks, firms can benefit from selecting solutions in wide use within the industry to get early exposure and response to new threats. While technology is a critical part of any cybersecurity program, the biggest weakness continues to be social engineering attacks that rely on the vulnerability of people. As a result, a comprehensive program needs to include active feedback, adaptive controls based upon who and what resources are targeted, and proactive training. As attackers learn about a firm's use of AI for different functions, it is likely that data manipulation attacks will start to emerge, aiming to bias data thereby influencing models. As a result, extra attention should be taken with regards to the protection of this data.

IX. Outsourcing and Vendor Management: Additional Considerations

Relating to *Outsourcing and Vendor Management* under the *Additional Considerations* section starting on page 19 of the FINRA Paper (PDF version), the Proofpoint CPC would like to add that AI models and the corresponding results are typically system-specific and require the underlying software to execute. When selecting a vendor, it is important to understand the solution's tracking of not only results on each piece of data but also how the system was configured over time. Most cloud solutions are updated on a regular basis. Reviewing the vendor's version management is also critical for model explainability as the behavior of a system at a given point in time requires access to the system (and/or the source code) in use at that point. Similarly, the financial viability of the vendor it a critical consideration, as the aforementioned access to historical data, systems and expertise is dependent on the firm being a going concern.

X. Books and Records Considerations Related to the Use of AI Technology

Finally, relating to *Books and Records* under the *Additional Considerations* section starting on page 19 of the FINRA Paper (PDF version), the CPC agrees that, "The use of AI applications may lead to the creation of new records" and that, "Firms should review the use of their AI tools and systems to ensure compliance with recordkeeping obligations, such as those associated with Exchange Act Rules 17a-3 and 17a-4 and FINRA Rule 4510 (Books and Records Requirements)." Further, Proofpoint and our Customer Product Council ("CPC") would like to open a broader discussion on Books and Records requirements related to the use of AI technology.

Proofpoint's technology provides electronic communications capture, storage, and monitoring for our CPC customers; the retention of which are required under SEA 17a-4 and FINRA Rule 4510. In working with our customers, Proofpoint anticipates a possible issue when AI is used to convert audio, such as phone conversations, to text.

SEA 17a-4 and FINRA Rule 4510 require retention of designated books and records. In FINRA's recent FAQ on *COVID-19 / Coronavirus* related to *Advertising Regulation*, when asked whether meetings with clients via a live video or audio-conferencing platform should be keep as records and potentially supervised, FINRA answered:

> Members must supervise registered representatives' live meetings with customers via video or audio-conferencing platforms in a manner reasonably designed to achieve compliance with applicable securities laws and regulation and FINRA rules.
>
> Unless required to record pursuant to FINRA Rule 3170 (Tape Recording of Registered Persons by Certain Firms) or otherwise, members generally are not required to record live video or audio-conferences with customers. However, if a registered representative during the video or audio conference uses the chat or instant messaging feature of the platform or presents slides or other written (including electronic) communications, the member must keep records of these written communications in accordance with Securities Exchange Act

> Rule 17a-4 and FINRA Rules 3110.09 (Supervision) and 4511 (General Requirements), and their content must be consistent with applicable standards such as FINRA Rule 2210 (Communications with the Public) and 3110(b) (Supervision). Depending on the nature and number of persons attending the video meeting, these written communications may be correspondence, retail communications or institutional communications, and must be supervised as such. See FINRA Rules 2210(b) and 3110(b)(4).
>
> Moreover, if a member chooses to record live video or audio conversations with customers, the member may be required to produce the recording in connection with a regulatory request. If a firm permits public appearances through video or audio-conferencing platforms, the member must ensure compliance with FINRA Rule 2210(f).

Excluding instances under FINRA Rule 3170 (Tape Recording of Registered Persons by Certain Firms), members are *generally* not required to record live video or audio with customers, but text forms of communications with customers, such as via chat, instant messaging, or slide presentations may both be required for retention and supervision. (Securities Exchange Act Rule 17a-4 and FINRA Rules 3110.09 (Supervision) and 4511 (General Requirements), FINRA Rule 2210 (Communications with the Public) and 3110(b) (Supervision). There seems to be a major difference in how verbal and visual communications are treated versus written where, once communications are in writing then retention and supervision requirements are triggered.

In the instance where a phone conversation, again *generally* not required to be retained, is converted to 'text' using AI, may this then cause retention and supervision obligations? If this is not the case, then would the text rendering of an audio/visual communication be classified as metadata not requiring retention and supervision? Should retention and/or supervision rules be amended or clarified given today's technology advances since the original written correspondence rules were in place when the primary medium of communication were letters and fax (pre-1990's)?

Conclusion

Again, we appreciate the offer and willingness of FINRA to consider our feedback. As FINRA moves this initiative forward, please feel free to reach out to Proofpoint at any time for clarification, or further input. Proofpoint continues to make significant investments in the application of AI and Machine Learning to the benefit of FINRA member firms and we look forward to ongoing collaboration with the regulatory bodies to bring about balanced market efficiencies.

Sincerely,

R. Darren Lee
Executive Vice-President & GM, Compliance & Digital Risk
Proofpoint