April 24, 2024

Christopher J. Kirkpatrick
Secretary, Commodities Futures Trading Commission
Three Lafayette Center
1155 21st Street NW
Washington, DC 20581

Delivered via online submission.

> Re: Theta Lake, Inc.'s Response to the Request for Comment on the Use of Artificial Intelligence in CFTC-Regulated Markets (CFTC Release No. 8553-24)

Dear Mr. Kirkpatrick:

Theta Lake, Inc. ("Theta Lake") respectfully submits this letter in response to the request for comment ("RFC") on the use of artificial intelligence in Commodities Futures Trading Commission (the "Commission," or "CFTC")-regulated markets.

In this letter, Theta Lake provides an overview of its platform, and responds to Questions 2, 7, 11, 15, and 18 posed in the RFC.

Our goal is that Theta Lake's response to the RFC helps to, in the words of Commissioner Kristin N. Johnson, "shape the development and deployment of AI in CFTC-regulated markets in a manner that harnesses AI's many promises, while responding to the many new challenges that will arise."[1]

I. Theta Lake Overview

Theta Lake provides a multi-patented, AI-enabled, Digital Communications Governance ("DCG") platform that assists CFTC-regulated market participants to meet obligations for the capture, retention, search, and proactive compliance of electronic communications ("e-comms").[2] Theta Lake has been named as a Sample Vendor in two Gartner® Hype Cycles™—for both Storage and Data Protection and Data Security—and currently addresses CFTC compliance for regulated firms.

Theta Lake integrates with over 100 communications platforms including Bloomberg, Cloud9, Microsoft Teams, NICE, Verint, Webex, WhatsApp, Zoom, and more, to enable compliance and security controls to support the borderless, cloud native, device agnostic modern workforce. Additionally, Theta Lake counts Cisco, RingCentral, Salesforce/Slack, Zoom, and senior executives from Microsoft as its strategic investors.

---

[1] Commissioner Kristin N. Johnson's Statement on the CFTC RFC on AI: Building a Regulatory Framework for AI in Financial Markets, January 25, 2024. Last visited April 24, 2024.

[2] Gartner's webpage "Market Guide for Digital Communications Governance" summarizes DCG as "[t]he quantity and scope of communication tools, such as Microsoft Teams, Zoom and WhatsApp, makes it more complex for organizations to enforce corporate governance and regulatory compliance use policies." See Gartner website. Last visited April 24, 2024.

Last month, Theta Lake and Zoom announced the release of the Zoom Compliance Manager powered by Theta Lake, an "all-in-one offering that provides archiving, eDiscovery, legal hold, and information protection capabilities to help organizations fulfill regulatory requirements and mitigate organizational communications compliance risks across the Zoom platform."[3]

These direct investments and seamless product integrations demonstrate a meaningful validation of our approach to DCG and set us apart in the market segment.

Ultimately, Theta Lake's unified capture, search, and analysis capabilities allow firms to deploy their communications platforms of choice to facilitate AI-driven compliance at scale.

II. How Theta Lake Helps CFTC Registrants Capture E-Comms to Meet Books and Records Requirements (Question 2(e))

From a foundational perspective, Theta Lake provides new and novel features that support the entire lifecycle of electronic communications compliance—from API-based integrations with e-comms platforms that capture interactions with full context and fidelity to AI-enabled analysis of communications to identify risk as well as the secure storage of conversations consistent with CFTC recordkeeping requirements.

Following the 2008 Dodd-Frank reforms, entities across the regulated CFTC spectrum–Swap Dealers, Major Swap Participants, Futures Commission Merchants, and others–are required to demonstrate robust electronic communications record keeping protocols per CFTC Regulations § 23.202(b)(1) and § 1.35(a)(iii), which require capture and retention of "all oral and written communications provided or received" among multiple additional daily trading records. The CFTC's recordkeeping mandates for e-comms have been emphasized in recent enforcement actions focused on off-channel communications compliance failures.

Considering the ever-expanding use of multi-modal collaboration tools like Microsoft Teams, RingCentral, Webex, and Zoom, which provide integrated meeting, voice, chat, and productivity features, the complexity associated with capturing the communications from these platforms has increased. This capture complexity coupled with heightened scrutiny and enforcement activity around off-channel communications means that CFTC-regulated firms must prioritize e-comms capture and oversight as critical components of compliance programs.

Theta Lake's approach to communications capture is unique as we ingest data using platform APIs, meaning that we plug directly into the stream of interactions to capture all of the rich media associated with conversations such as chats, audio, links, emojis, reactions, file transfers, polling, Q&A, message modifications (i.e., edits and deletes), and more. Communications captured by Theta Lake are preserved in a manner that mirrors the look and feel of the conversation on the source platform, so that subsequent reviews benefit from a consistent and enriched look and feel.

---

[3] See Theta Lake Press Release dated March 20, 2024. Last visited April 21, 2024.

From a retention perspective, Theta Lake's platform meets the technical requirements of CFTC Regulation § 1.31 and offers additional features to enhance security protocols. Audit logging is enabled for all communications to demonstrate the "authenticity and reliability" of data capture, retention, alteration, and deletion events.[4] Data is secured in Theta Lake's SOC 2, Type II, PCI DSS, and TruSight audited platform to maintain the "security, signature, and data as necessary to ensure the authenticity of the information."[5] Customers may set retention periods in the platform to meet CFTC requirements and, in fact, may set any number of desired retention periods to accommodate other regulatory or operational data governance needs, including those that may be necessary if an organization is registered under multiple regulatory regimes. Unified search allows organizations to produce records efficiently and effectively, and our cloud-based resiliency and redundancy controls support ongoing availability of content.[6]

In addition to technical controls that support compliance with Regulation § 1.31, Theta Lake customers may bring their own encryption keys to the platform and bring their own cloud storage. So, for example, a customer using Amazon Web Services or Microsoft Azure may use Theta Lake to capture communications from various e-comms platforms and route data to their own cloud storage for long-term retention. These bring your own encryption and storage capabilities boost the already strong cybersecurity protections offered in Theta Lake's platform and facilitate end-to-end management of communications data.

III.   How Theta Lake Uses AI to Enhance CFTC Compliance (Question 2(d))

Theta Lake deploys AI in its platform to support firms' compliance with CFTC requirements related to the capture, retention, and oversight of e-comms. Theta Lake's AI-powered detections identify risk in captured communications, large language models support conversation summary features, and Theta Lake takes a user-centered approach to AI that incorporates a human-in-the-loop. We discuss these components in more detail below.

Theta Lake's AI-Powered Risk Detections

Following ingestion of e-comms content, firms can use Theta Lake's AI capabilities to analyze conversions for risk. Theta Lake's AI-powered models can detect myriad risks including conversations indicating potential off-channel activity, communications about derivatives products at the core of the CFTC's daily trading records rules as well as conversations that may be indicative of collusion, coercion, or promissory statements like guarantees. Risk detections also extend into categories that determine if an AI notetaker bot or meeting assistant is present in a meeting, or whether a disclaimer or disclosure is provided during a meeting. Risk detections can be applied across any media type, meaning detections will trigger if a risk appears in a recorded of a telephone conversation of an associated person; in the video of a screen share, whiteboard, or webcam; or in a chat, email, or its corresponding attachments, emojis, or reactions.

---

[4] CFTC Regulation 1.31(c)(1).
[5] CFTC Regulation 1.31.(c)(2)(i).
[6] CFTC Regulation 1.31.(c)(2)(ii).

Theta Lake's AI-based risk detections improve on legacy approaches to analysis, which are rooted in lexicon-based searches that simply attempt to match the appearance of specific words or terms, regardless of context, as indicative of the presence of risk. These lexicon-based approaches are inflexible and limiting in contrast to those leveraging AI, which can generalize to patterns beyond simplistic words or phrases.

AI enables Theta Lake's risk detections to be resilient to errors like misspelled words, transcripts that contain inaccuracies ("WhatsApp" mis-transcribed as "WhatsUp"); or incorrect data ingested from the application of Optical Character Recognition to video or images ("MS Word" scanned as "M5 W0rd"). Theta Lake's elasticity to these nuanced errors and imperfections cannot be matched by antiquated, lexicon-based supervision.

Theta Lake's machine learning models interpret the context of a conversation to better differentiate between true and false positives.

If a firm supervises oral communications using a lexicon-based system that scans audio transcripts, the system will fail if it cannot determine the context of conversation or disambiguate content to overcome spelling or transcription errors.

If two derivatives traders are discussing a "vanilla rate swap" on a recorded telephone line and that phrase is mis-transcribed as "Fanella rates what," Theta Lake's machine learning model for detecting derivatives conversations will identify the phrase as Dodd-Frank relevant, despite the fractured, error prone transcription.

Moreover, the comprehensive nature of Theta Lake's e-comms capture capability reinforces and improves the application of its AI-based risk detections. Theta Lake's machine learning models can be combined with detections for emojis, reactions, or file types to refine and improve a firm's ability to identify conduct risk. So, the detection for off-channel conversations can be combined with a detection for the mobile phone emoji (📱), or the collusion detection can be coupled with the rocket ship or moneybags emojis (🚀 or 💰) to better surface conversations that are unique to an organization's business risks.

Risk detections can be further modulated to apply to specific groups, geographies, or communications platforms to uplevel compliance supervisory practices. A rule could be created to apply only to employees on the Chicago commodities desk using Zoom Chat, which triggers when the off-channel risk detection triggers and where the mobile phone emoji (📱) appears.

These granular and dynamic detection rules can only be deployed if the underlying communications are captured in a comprehensive manner that preserves elements like videos, voice, images, emojis, and reactions to enable subsequent analysis by machine learning algorithms.

AI Conversation Summaries

Theta Lake employs AI, using large language models, to support its conversation summary feature. Conversation summaries allow organizations to condense weeks- or months-long chats and multi-

platform conversations into concise summaries, making it easier for compliance teams to review and understand the essence of an interaction. By analyzing the content of conversations across various communication modalities, including video, audio, and chat, Theta Lake generates summaries that capture key conversation details including themes, participants, and entities.

Summarization is particularly valuable in supervision processes, where large volumes of communication data must be analyzed efficiently to expedite compliance reviews, prepare for regulatory inquiries, or spot check data prior to production to outside counsel.

<u>Human-in-the-Loop</u>

Theta Lake adopts a human-in-the-loop approach to the use of AI in its risk detections, which enhances the rigor of oversight processes. When conversations trigger risks, they are presented to human reviewers for confirmation and analysis. Theta Lake visually pinpoints where its AI models have identified a risk by labeling or highlighting the relevant section of a conversation, so that reviewers can quickly spot and analyze key portions of the interaction.

This human review step allows for the application of a compliance team's expert judgment and context-specific understanding to validate AI-identified risks.

IV.   Theta Lake's Approach to AI Governance (Question 7)

With respect to governance, Theta Lake's data science team is staffed by scientists and engineers with distinguished academic backgrounds and decades of collective experience building machine learning models for financial services and other regulated industries. The team has developed over 100 machine learning models, which are pre-trained and ready for use to help customers scan multiple media types for relevant risks.

Theta Lake adheres to a rigorous Software Development Lifecycle ("SDLC") process as outlined in its Information Security Policy. The SDLC ensures that the development and deployment of AI models are carried out in a secure and controlled manner where routine testing for potential vulnerabilities and other issues are completed prior to deployment.

Continuous improvement and ongoing oversight are also key to Theta Lake's AI governance. Theta Lake continuously refines its models to enhance accuracy, improve performance, and manage scope. By regularly updating models, Theta Lake ensures that models remain current and capable of addressing evolving regulatory compliance expectations and other changes in DCG-related risks.

Theta Lake focuses on the long-term performance of its models to minimize data and concept drift. Models are monitored post-deployment for unexpected changes and to incorporate feedback from customers around false positives and negatives. Models are refined and retrained routinely over time, incorporating new data sources in training data sets to account for emerging patterns.

A great example of Theta Lake's continuous model refinement is evidenced in the evolution of its off-channel communications risk detection. The base detection for off-channel conversations has been developed and enhanced over time, incorporating insights from recent enforcement actions and the ability to detect relevant discussions in foreign languages. This ongoing refinement enables Theta Lake and its customers to stay ahead of emerging risks and ensure that AI models can effectively identify potential compliance issues.

V.   Theta Lake's Platform Features for Explainability and Transparency (Question 11)

Foundational to its transparency practices, Theta Lake provides customers with full control over how AI detections are enabled and operate within the platform. Customers choose which specific detections are enabled and can activate or disable them anytime thereafter. Customers manage the risk sensitivity of detections, categorizing them as risky, informational, or validations, which result in higher, lower, or neutral aggregated conversation risk scores. Risk scores allow compliance teams to prioritize how conversations are reviewed—high risk scores drive near-time reviews of key conversations, whereas low- or no-risk interactions can be sampled on a percentage basis for spot checking purposes. The fine-grained transparency controls in Theta Lake's platform allow organizations to customize AI behavior and performance to address narrowly tailored compliance requirements.

Additionally, Theta Lake provides reporting on AI transparency and performance metrics as well as mechanisms for providing feedback about potentially anomalous results.

Theta Lake's classifier audit report provides metrics around hit rates for a given risk detection over time. So, for example, customers can query the number of Zoom video meetings triggering the detection for misleading or inappropriate promotions across a specific time range and group of users.

Customers can also provide direct feedback to Theta Lake if models generate false positives or negatives to improve performance and account for evolving communication characteristics.

VI.   How Theta Lake Mitigates the Risk of Bias (Question 15)

Theta Lake's AI is designed to be fair and objective in its risk detection capabilities. The AI models used by Theta Lake are trained on thousands of positive and negative examples of the risks they aim to detect, without incorporating characteristics that may introduce bias, such as age, race, religion, sexual orientation, or ethnicity. This means that the AI-enabled detections are focused solely on the contextual meaning of a conversation and the potential risks involved. Theta Lake's risk detections are not developed for high-risk purposes like underwriting, determining capital and reserve adequacy, investment advice, consumer lending, banking, or credit determinations.

VII.   Theta Lake's Security and Audit Processes Support AI Development and Third-Party Risk Assessment (Question 18)

As a third-party service provider to enterprise financial services organizations, Theta Lake understands the primacy of robust, auditable security protocols, many of which directly support AI development.

Core audit standards like SOC 2, Type II, PCI DSS, TruSight, and ISO 27001 sit at the intersection of third-party risk and cybersecurity. These third-party audits provide consistent, repeatable, measurable protocols that demonstrate compliance with articulated controls, including those critical to AI development processes.

Third parties providing AI services to CFTC-regulated firms should rely on these audit standards to evidence baseline controls. Likewise, firms conducting vendor due diligence should leverage these audit reports as an essential component of risk assessment processes. These audits collect a set of uniform controls and apply them to companies of all shapes and sizes, offering a standardized metric with which to assess third parties for whom AI is a core service offering.

VIII.   Conclusion

We hope that this background about how Theta Lake uses AI to support record keeping and compliance as well as additional background on governance, transparency, and third-party risk assists the CFTC as it examines the broad adoption of this new technology. Theta Lake is in regular dialogue with US and international regulators and would welcome the opportunity to discuss these issues with the CFTC to provide additional context and demonstrate the value of AI for these evolving use cases.

Respectfully submitted,

/s/ Marc Gilman
Marc Gilman
General Counsel and VP of Compliance

/s/ Sharon Hüffner
Sharon Hüffner
Chief Scientist

/s/ Rohit Jain
Rohit Jain
Distinguished Engineer

/s/ Stacey English
Stacey English
Director of Market Intelligence

/s/ Susannah Hammond
Susannah Hammond
Senior Regulatory Intelligence Expert