Part 2–Helpful Guidance on Managing (Non-Cybersecurity) AI Risks from Hong Kong's SFC

November 22, 2024

In Part 1 of this series, we discussed the recent Circular and accompanying Appendix issued by Hong Kong's Security and Futures Commission (the "SFC") on cybersecurity risks and mitigations related to the use of generative artificial intelligence language models ("AI LMs"). In this Part 2, we discuss the SFC's expectations for how licensed corporations ("LCs") (generally securities and futures markets participants such as private equity firms, asset managers or hedge funds that are licensed by the SFC to carry out regulated activities) should be managing and mitigating non-cybersecurity risks, including through AI governance and model risk management.

As we indicated in Part 1, due to the clear and practical guidance offered by the Circular, which is effective immediately, we suspect it will be influential in how other financial regulators address AI-related risks, and it therefore can serve as a valuable AI risk assessment framework for financial firms that are not LCs.

Scope of the Circular

As we explained in detail in Part 1, the Circular applies to LCs in relation to their regulated activities (*e.g.*, advising on securities, automated trading services, and asset management). The Circular's requirements are risk-based and should be implemented "commensurate with the materiality of the impact and the level of risk presented by the specific use case or application of the AI LM."

Non-Cybersecurity Risks Associated with AI LMs

The Circular and Appendix identify a non-exhaustive list of AI risks to which AI LMs are susceptible and which may, if unmitigated, lead to potential client or investor harm. These include:

- **Hallucination Risk**: AI LMs are prone to "hallucinations," *i.e.*, providing responses that while factually wrong might appear plausible to a user. The Circular notes that when adopting an AI solution marketed as eliminating hallucination, LCs should assess the solution's reliability, since such offerings are found to have limitations.

- **Bias Risk**: Biased input can lead to biased outputs. This bias can be introduced in a variety of ways (*e.g.*, through the data used to train the LM, developer assumptions, model design, and implementation choices).

- **Drift**: AI LMs may degrade or "drift" over time such that they no longer do what they could do originally.

To address these and related risks, the Circular requires mitigation in the form of senior management oversight, model risk management, and third-party risk management.

Senior Management Responsibilities

The Circular makes clear that LCs should have the resources and procedures needed for the proper performance of its business activities. For LMs, that means that senior management should ensure that, throughout the lifecycle of an LM, the LC has (1) implemented effective policies, procedures and internal controls, and (2) adequate senior management oversight and governance by suitably qualified and experienced individuals. The model lifecycle covers model development (*i.e.*, design, implementation, customization, training, testing and calibration) and model management (*i.e.*, validation, approval, ongoing review and monitoring, use and decommissioning). Other responsibilities of senior management include:

- **Identifying of High Risks**: The governance framework should address the identification of high-risk use cases, taking into consideration any potential adverse client impact, particularly if the AI LM's output is inaccurate or inappropriate. While the SFC does not provide an exhaustive list of what it deems "high-risk," the Circular states that "[g]enerally speaking the SFC considers using an AI LM for providing investment recommendations, investment advice or investment research to investors or clients as high-risk use cases" for which extra risk mitigation measures should be adopted (discussed below).

- **Cross-Functional Approach**: Senior management should ensure that responsible staff from the business, risk, compliance and technology functions can effectively manage the LC's adoption and implementation of AI LMs by possessing the relevant competence in AI, data science, model risk management, and domain expertise. The

legal and compliance function should assess the use of AI LMs from a compliance risk perspective, including whether their deployment may undermine the LC's compliance with applicable legal and regulatory requirements.

- **Model Risk Awareness**: The LC and its senior management should ensure that they are aware of the risks and limitations of an AI LM and the input data, and that the AI LM deployed is fit for purpose and appropriate for the specific use case, given those risks and limitations.

- **Maintaining Ultimate Responsibility**: While delegation of certain functions to group companies is permitted, the LC remains responsible for ensuring its compliance with applicable legal and regulatory requirements. If the delegated function relates to the use of AI LMs in a high-risk use case, the LC should also ensure it has sufficient management oversight and ongoing monitoring of its deployment of the AI LMs.

AI Model Risk Management

The Circular requires LCs to maintain an effective AI model risk management framework, which includes:

- **Risk-Based Mitigation**: LCs should take risk mitigation measures commensurate with the materiality of the impact and risks of the specific use case, particularly to address the LM's hallucination risk. LCs remain accountable for their output regardless of the risk mitigation measures adopted.

- **Segregation of Responsibility**: If an LC undertakes model development activities, the LC should ensure that the model development function is segregated from other functions responsible for model validation, approval and ongoing review and monitoring, where practicable and considering the use case and the level of risk involved.

- **Model Validation**: LCs should subject LMs to proper validation measures before approving them for use, as well as when any material changes are made to its design, assumptions, inputs, calculations or outputs. The scope of model validation should cover testing the effectiveness of the cybersecurity and data risk management controls.

- **Model Testing**: LCs should assess model performance by conducting comprehensive end-to-end testing that covers the entire processes from user input to system output

including all related system components or functionalities, such as retrieval augmented generation ("RAG"), content filtering, or prompt management solutions.

- **Ongoing Review & Monitoring**: LCs should subject the performance of AI LMs to ongoing review and monitoring to ensure that they remain fit for purpose and continue to function as intended, particularly after events such as changes in the underlying market dynamics or economic regime, or the inclusion of a new dataset by the LC to fine-tune the LM.

- **User Disclosures**: If an LM is used in the LC's client interface, the LC should provide prominent disclosures that the user is interacting with AI rather than humans, and that the output generated by the LM may not be accurate.

- **Data Quality**: LCs are expected to also ensure the quality of the data used to train an LM, including identifying and mitigating biases that may have had a material impact on the LC's use cases.

- **Documentation**: The results of any model testing, calibration, validation, and ongoing review should be documented.

For high-risk use cases, LCs are further required to:

- **Monitor Model Accuracy**: Conduct model validation, ongoing review, and monitoring in relation to the performance of the LM to improve factual accuracy to a level commensurate with the specific use case.

- **Human-in-the-Loop**: Have a human in the loop to address hallucination risk and review the AI LM's output for factual accuracy before relaying it to the user.

- **Consistency Testing**: Test output robustness to prompt variations, as it has been reported that AI LMs may generate different predictions based on text inputs that have the same meaning.

- **Regular Disclosures**: In the case of an LM used in a client interface, users should be provided with the disclosure that (1) they are interacting with AI rather than humans, and (2) the output generated by the AI LM may not be accurate every time they interact with the AI LM (rather than a one-off disclosure).

- **Ongoing Monitoring**: Because new risks may emerge, LCs should continue to test and monitor their LMs for high-risk use cases, even if a human is reviewing the output after deployment.

Third-Party Provider Risk Management

In the case of AI LMs provided by a third-party provider ("Provider"), the SFC also requires:

- **Model Validation**: When performing model validation on a Provider's LM with limited transparency or information on hand, the LC should assess:
  - to the extent practicable, whether the Provider itself has an effective model risk management framework; and
  - whether the output and performance of the AI LM are appropriate for the LC's specific use cases, including considering the model risk management framework with respect to its use cases and adopting risk mitigation measures as appropriate.
- **Open Source Risk Mitigation**: Where an open-source AI LM is not provided by an identifiable Provider or it is not practicable to apply the third-party risk management requirements (such as performing due diligence or ongoing monitoring), an LC should nevertheless ensure that the open source AI LM is subject to other applicable requirements, including the firm's relevant model development and model management measures discussed above.

Notification Requirements

LCs that intend to use AI LMs in high-risk use cases should comply with the notification requirements laid out under the SFC's Information Rules, which require intermediaries to notify the SFC of any significant changes in the nature of their business and the types of service they provided. The SFC encourages LCs deploying high-risk AI to also discuss their plans with the SFC as early as possible, preferably at the business planning and development stage, to avoid potential adverse regulatory implications.

Practical Takeaways

Gap Assessment and Budget

As we stated in Part 1, LCs should consider conducting a risk assessment that identifies gaps between the requirements in the Circular and their own AI compliance programs and building a road map for closing any material gaps. For some firms, it may take

significant time and resources to fully implement these new requirements, and so they may want to start early. Even firms that are not subject to the Circular may consider conducting a gap analysis in anticipation that similar rules are likely to be adopted by other regulators and may be considered best practices in AI governance and risk management. For some firms, compliance with the Circular will require a significant increase in their compliance budgets and the securing of additional resources for 2025 and beyond.

Training

The Circular does not explicitly mention training as a requirement, but in light of new obligations set forth in the Circular, and in practice, the number of responsibilities that rest with senior management, LCs should consider providing training on AI-risk management generally and compliance with the Circular in particular.

The authors would like to thank Debevoise Law Clerks Adam Shankman and Diane Bernabei for their contribution to this blog post.

* * *

Please do not hesitate to contact us with any questions.

To subscribe to the Data Blog, please click here.

The Debevoise Data Portal is an online suite of tools that help our clients quickly assess their federal, state, and international breach notification and substantive cybersecurity obligations. Please contact us at dataportal@debevoise.com for more information.

The cover art used in this blog post was generated by DALL-E.

Luke Dembosky
Partner, Washington, D.C.
+ 1 202 383 8020
ldembosky@debevoise.com

Avi Gesser
Partner, New York
+ 1 212 909 6577
agesser@debevoise.com

Gareth Hughes
Partner, Hong Kong
+ 852 2160 9808
ghughes@debevoise.com

Erez Liebermann
Partner, New York
+ 1 212 909 6224
eliebermann@debevoise.com

Matt Kelly
Counsel, New York
+ 1 212 909 6990
makelly@debevoise.com

Emily Lam
International Counsel,
Hong Kong
+ 852 2160 9823
elam@debevoise.com

Suchita Mandavilli Brundage
Associate, New York
+ 1 212 909 6483
smbrundage@debevoise.com