12 November 2024

## Circular to licensed corporations

## Use of generative AI language models

1. With the introduction of generative artificial intelligence language models (AI LMs) into the public domain, both commercial and open source AI LMs are now readily accessible to financial institutions. The use of AI LMs may enable licensed corporations (LCs) to handle client interactions as well as internal manual processes and operations more efficiently, thereby freeing up manpower for other value-adding tasks and improving overall productivity.

2. Based on the Securities and Futures Commission's (SFC) engagement exercise with a cross section of international and local LCs, the SFC notes that firms are leveraging AI LMs to respond to client enquiries via public facing chatbots, summarize information, generate research reports, identify investment signals as part of the investment decision making process, or generate computer code during the development of software applications.

3. The SFC encourages and supports the responsible use of AI and AI LMs by LCs to innovate, deliver products or services more effectively or enhance their operational efficiency. While traditional AI has been widely adopted by financial institutions for decades, AI LMs may amplify existing risks and pose additional risks on top of those from traditional AI. AI LMs democratize access to AI as they take natural language instructions from users as input such that very little technical proficiency is required to use them. The lower entry barriers for firms without the technical expertise in traditional AI to use AI LMs may result in firms deploying such technology before proper risk mitigation measures are put in place. Furthermore, the ability of AI LMs to output human-like responses may result in over-reliance, with users accepting their outputs without critical evaluation.

### Risks in relation to AI LMs

4. AI LMs are susceptible to the following risks. If not managed properly, the following risks could have negative legal, reputational, operational or financial impacts on LCs, which in turn may harm clients or investors:

    (a) AI LMs' output can be inaccurate, biased, unreliable and inconsistent. For instance:

        (i) AI LMs are prone to hallucination risk, ie, providing plausible responses to enquiries which are in fact wrong, including systematically echoing the user's[1] opinions regardless of the accuracy of the user's statement;

[1] The term "user" in this circular may refer to a member of the LC's staff, its client or another entity (which is not necessarily the LC's client) making use of its AI LM, and should be construed in accordance with the actual circumstances of the use case.

(ii) Biases may exist in the data used to train AI LMs, in the input representation (when data is transformed into numerical input to feed into the model), and in the model developer's assumptions, model design and implementation choices, which may result in biased, inappropriate or discriminatory outputs; and

(iii) An AI LM's performance may drift and degrade over time such that it no longer does what it was initially designed to do.

(b) There are heightened risks of cyberattacks, inadvertent leakage of confidential information in relation to a firm or its clients, as well as breaches of personal data privacy and intellectual property laws.

(c) Firms may be reliant on external service providers to develop, train and maintain the AI LMs. Given the limited number of such external service providers, firms are exposed to the risks of concentration and operational resilience in the event of system unavailability.

5. To facilitate the industry's responsible adoption of AI LMs, this circular sets out the SFC's expectations on LCs in relation to their use. LCs should consider all risk factors relevant to their particular AI LM use cases and implement risk mitigation measures as appropriate. The Appendix sets out a list of non-exhaustive risk factors for LCs' reference. As this field is fast moving, if necessary, the SFC will engage with the industry to develop more specific guidance in relation to managing those risks, as well as consider how to facilitate financial firms' capacity building in relation to AI LMs.

**Scope of this circular**

6. The requirements of this circular apply to LCs offering services or functionality provided by AI LMs or AI LM-based third party products in relation to their regulated activities[2]. This circular is applicable regardless of whether the AI LM is developed or provided by the LC itself, its group company, an external service provider (Third Party Provider) or comes from an open source.

**Risk-based approach**

7. An LC may implement the requirements in this circular, including the Core Principles detailed below, in a risk-based manner, commensurate with the materiality of the impact and the level of risk presented by the specific use case or application of the AI LM.

8. Generally speaking, the SFC considers using an AI LM for providing investment recommendations, investment advice or investment research to investors or clients[3] as high-risk use cases, given that problematic output from the AI LM may lead LCs to recommend unsuitable financial products to their clients or misinform investors in their decision making. LCs should adopt extra risk mitigation measures for high-risk use cases (see paragraphs 18 – 19).

[2] Including "relevant activities" with respect to virtual asset trading platform operators.
[3] For the avoidance of doubt, this does not encompass after sales client servicing.

## (A) Core Principle 1: Senior management responsibilities

9. An LC should have the resources and procedures needed for the proper performance of its business activities[4]. An LC's senior management[5] should ensure that, throughout the full lifecycle of an AI LM:

   (a) Effective policies, procedures and internal controls[6] are implemented; and

   (b) Adequate senior management oversight and governance by suitably qualified and experienced individuals are in place[7].

   The model lifecycle covers Model Development (ie design, implementation, customisation, training, testing and calibration) and Model Management (ie validation, approval, ongoing review and monitoring, use and decommissioning).

   The governance framework should encompass the identification of high-risk use cases by taking into consideration any potential adverse client impact, particularly if the AI LM's output is inaccurate or inappropriate.

10. Since the oversight and risk management of AI LMs should be performed by fit and proper staff[8], the LC's 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.

11. To properly manage the use of AI LMs, 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[9].

12. Whilst an LC may delegate to its group company certain functions, such as the performance of model validation, it remains responsible for ensuring its compliance with the 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.

[4] General principle 3 of the Code of Conduct for Persons Licensed by or Registered with the Securities and Futures Commission (Code of Conduct).
[5] General principle 9 of the Code of Conduct.
[6] Paragraph I(1) of the Management, Supervision and Internal Control Guidelines for Persons Licensed by or Registered with the Securities and Futures Commission (Internal Control Guidelines).
[7] Paragraph I(5) of the Internal Control Guidelines.
[8] Paragraph 4.1 of the Code of Conduct.
[9] Paragraph 14.1 of the Code of Conduct.

## (B) Core Principle 2: AI model risk management

13. As part of an effective AI model risk management framework[10], an LC should:

    (a) if it undertakes Model Development activities, have a Model Development function which is segregated[11] from the function which performs model validation, approval and ongoing review and monitoring, where practicable and having regard to the use case and the level of risk involved;

    (b) subject AI LMs to adequate validation to address any issues[12] (i) prior to approving them for use, and (ii) when material changes are made to its design, assumptions, input, calculations or output; the scope of model validation should cover testing the effectiveness of the cybersecurity and data risk management controls in relation to the AI LM[13];

    (c) assess model performance by conducting comprehensive end-to-end testing which covers the entire process 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; and

    (d) 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[14], 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 AI LM.

    The results of the model testing and calibration (to the extent that the LC carries out such activities), validation and ongoing review and monitoring should be documented.

14. The Model Development requirements apply only if the LC undertakes activities to develop, customise, refine or enhance an AI LM, such as fine-tuning, applying RAG or content filtering, or integrating external tools (such as prompt management solutions) with a pre-trained AI LM developed by a Third Party Provider.

15. The Model Development requirements do not apply if an LC (a) uses an AI LM (or an AI LM-based product) off-the-shelf and merely configures essential parameters such as the temperature, freezes the underlying AI LM without further development or customisation, or provides disclosures to the user in the AI LM user interface; or (b) integrates an off-the-shelf product with an AI LM without customisation in other components of an AI LM system architecture. These products should nevertheless be subject to proper Model Management.

[10] The objective of paragraph VIII of the Internal Control Guidelines.
[11] Paragraph II of the Internal Control Guidelines.
[12] Paragraph IV(5) of the Internal Control Guidelines.
[13] It is not the SFC's requirement that LCs' AI model risk management framework duplicate the firms' existing frameworks in relation to cybersecurity, data and Third Party Provider risk management. It would suffice as long as the LC's enterprise-wide cybersecurity, data and Third Party Provider risk management framework covers the requirements of this circular.
[14] Paragraph IV(4) of the Internal Control Guidelines. LCs should beware that merely reviewing industry standard benchmark tests on the AI LM's performance may not be sufficient.

*Risk mitigation measures – general*

16. LCs should take risk mitigation measures commensurate with the materiality of the impact and risks of the specific use case, particularly to address the AI LM's hallucination risk. LCs adopting solutions marketed as eliminating or avoiding hallucination should thoroughly assess their reliability, since such offerings are found to have limitations. LCs remain accountable for their output regardless of the risk mitigation measures adopted.

17. Where an AI LM is used in the LC's client interface, the LC should provide prominent disclosures in the user interface that they are interacting with AI rather than humans and that the output generated by the AI LM may not be accurate[15].

*Risk mitigation measures - high-risk use cases*

18. For high-risk use cases, LCs should adopt risk mitigation measures including:

    (a) conducting model validation, ongoing review and monitoring in relation to the performance of the AI LM so as to improve its factual accuracy to a level commensurate with the specific use case;

    (b) having 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[16];

    (c) testing 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; and

    (d) making the disclosures mentioned in paragraph 17 whenever the client interacts with the AI LM (as opposed to making a one-off disclosure upfront).

19. New properties, capabilities, behaviours and therefore risks of AI LMs may emerge given the fast-evolving technology landscape and the adoption of newer, upgraded models. As such, it is critical that LCs continue to test and monitor their AI LMs for high-risk use cases, even though a human in the loop reviews the AI LMs' output after deployment.

**(C) Core Principle 3: Cybersecurity and data risk management**

20. LCs should keep abreast of the current and emerging cybersecurity threat landscape[17] in relation to AI LMs and have effective policies, procedures and internal controls in place to manage the associated cybersecurity risks[18], including measures to promptly identify cybersecurity intrusions and, where appropriate, suspend the use of an AI LM.

21. In particular, adversarial attacks can steal or infer confidential information from an AI LM's training data, trick an AI LM into outputting incorrect or misaligned responses,

[15] General principle 5 of the Code of Conduct.
[16] Depending on the specific circumstances of the high-risk use case, the SFC will consider providing flexibility to LCs in the implementation of this requirement.
[17] See for example *Adversarial Machine Learning, A taxonomy and Terminology of Attacks and Mitigations* by National Institute of Standards and Technology, January 2024, and OWASP *LLM AI Cybersecurity & Governance Checklist*.
[18] Paragraph IV(2) of the Internal Control Guidelines.

override system prompts, or run malicious codes remotely. As such, LCs' cybersecurity measures should encompass adversarial attacks against the AI LM as well as the data used to train or fine-tune it. LCs should conduct adversarial testing periodically, to the extent practicable, on AI LMs to harden and protect them against adversarial attacks.

22. LCs should encrypt non-public data at rest and in transit to ensure their confidentiality and security[19]. LCs should note that the use of AI LM-based browser extensions may entail privacy and data leakage risks. LCs should therefore mitigate risks as appropriate, especially if staff have ready access to browser extensions.

23. In addition to the requirements in the circular on data risk management, the SFC expects LCs to ensure the quality of the data used to train an AI LM, including identifying and mitigating biases which may have a material impact on the LCs' use cases. LCs should also have due regard for the Artificial Intelligence: Model Personal Data Protection Framework by the Office of the Privacy Commissioner for Personal Data.

24. Given that training data extraction attacks exploit the ability of AI LMs to memorise and output sequences from their training dataset, LCs should have controls to assess and mitigate the risks of sensitive confidential information, such as personal data, being input by users or fed into the AI LM.

25. The LC should ensure that controls in relation to confidential client and business information remain effective throughout the model lifecycle[20].

**(D) Core Principle 4: Third Party Provider risk management**

26. An LC should exercise due skill, care and diligence in its selection of a Third Party Provider, including performing appropriate due diligence and ongoing monitoring to assess whether the Third Party Provider possesses the requisite skills, expertise, resources and controls to deliver the product or service to standards acceptable to the LC. In particular:

    (a) When performing model validation on a Third Party Provider's AI LM with limited transparency or information on hand, the LC should assess (i) to the extent practicable, whether the Third Party Provider itself has an effective model risk management framework, and (ii) whether the output and performance of the AI LM are appropriate for the LC's specific use cases, including considering the model risk with respect to its use cases and adopting risk mitigation measures as appropriate[21];

[19] Note that researchers have identified the possibility of an AI LM side channel attack.
[20] For example, LCs should consider the need for data segregation and access controls not only for training data but also for services that store or process embeddings and vectors, and whether information confidential to a specific business / function can be commingled with other information when training the model for different use cases across multiple businesses or functions within the organisation.
[21] If an LC is unable to perform sufficient due diligence to ascertain the robustness of the Third Party Provider's model risk management framework, an LC should take this fact into consideration when implementing its risk mitigation measures, including, for example, the frequency and depth of ongoing review and monitoring on model performance. Whilst the results of some industry standard benchmark tests of Third Party Providers' pre-trained AI LMs are available on the internet, LCs are reminded to ensure that the AI LM deployed is fit for purpose for their specific use cases, taking into account any Model Development activities undertaken by the LCs on top of the pre-trained AI LM.

(b) Where an open source AI LM is not provided by an identifiable Third Party Provider or it is not practicable to apply the Third Party Provider risk management requirements (such as performing due diligence or ongoing monitoring on the Third Party Provider), an LC should nevertheless ensure that the open source AI LM is subject to the other applicable requirements, including the firm's relevant Model Development and Model Management measures referred to in paragraph 13; and

(c) With respect to data management, the LC should assess if a breach by the Third Party Provider of applicable personal data privacy or intellectual property laws[22] could have a material adverse impact on the LC or its use cases, and whether the Third Party Provider has measures in place to protect or indemnify the LC against legal actions or claims against the LC in relation to the LC's use of the AI LM in case of any alleged breach of such laws.

27. An LC using an AI LM from a Third Party Provider should ensure that the allocation of responsibilities between itself and the Third Party Provider in relation to managing cybersecurity risks are well-defined and clearly understood.

28. Where the LC's development and deployment of Third Party Providers' AI LMs are undertaken with the use of Third Party Providers' data or software, including embedding models, vector stores, prompt management solutions, orchestration tools or performance evaluation tools, the LC should assess supply chain vulnerabilities as well as data leakage risk at each third party component of the LC's AI LM architecture, and apply stringent cybersecurity controls. An inventory of Third Party Providers' software should be maintained for cybersecurity monitoring.

29. LCs using Third Party Providers' AI LMs should assess their level of dependence on the prompt and consistent delivery and availability of services by the Third Party Providers, as well as the potential operational impact on them and their clients if the services are disrupted. LCs should establish appropriate contingency plans to ensure their operational resilience, particularly in relation to critical operations, if the use of AI LMs is disrupted or suspended.

**Notification requirements**

30. For LCs which intend to adopt AI LMs in high-risk use cases, they are reminded to comply with the notification requirements under the Securities and Futures (Licensing and Registration) (Information) Rules (Information Rules). These require intermediaries to notify the SFC of any significant changes in the nature of their business and the types of service they provide[23]. Moreover, they are encouraged to discuss their plans with the SFC as early as possible, preferably at the business planning and development stage, to avoid potential adverse regulatory implications.

[22] LCs should have regard to paragraph 12.1 of the Code of Conduct.
[23] Section 4 and Schedule 3 to the Information Rules. Please also refer to our circular dated 11 May 2015 entitled "Circular to Intermediaries Regarding Compliance with Notification Requirements".

31. This circular takes immediate effect. LCs should critically review their existing policies, procedures and internal controls to ensure proper implementation of, and full compliance with, the requirements in this circular. Nevertheless, the SFC recognises that some LCs may need time to update their policies and procedures to meet these requirements and the SFC will take a pragmatic approach in assessing LCs' compliance with the circular.

32. Should you have any queries regarding this circular, please contact your case officers-in-charge.

Intermediaries Supervision Department
Intermediaries Division
Securities and Futures Commission

End

SFO/IS/036/2024