

人工智能應用風險合規管理辦法 (摘要)

目的

本管理辦法旨在為本集團內部負責任且合乎道德地使用人工智能建立一個框架，以風險為本的方式開發、採購、實施及使用人工智能。考慮到人工智能的快速發展及其對金融服務業的影響，本管理辦法將因應合規、行業標準和道德考量而適時作出修訂。

人工智能應用的風險管理原則

1. 足夠的專業知識及培訓

本集團投入適當的資源以持續支持和發展人工智能應用。

2. 人工智能應用的可解釋性

本集團按人工智能應用的重要性，採取相稱的措施確保人工智能應用能提供合適程度的解釋，令相關持份者能基本明白人工智能應用的決策及輸出結果。

3. 高質量的數據管理

本集團使用高質量數據，並對數據質量進行檢查及監控，以滿足數據質量要求及數據質量評估準則及其他人工智能相關的數據準則。

4. 嚴格評估核實

本集團在人工智能應用投產前均需完成評估核實，包括評估其準確率、可靠性、穩健性及公平性，確保人工智能模型按預期運作才可投產。

5. 人工智能應用的可審計性

本集團為人工智能應用建立及保留足夠的稽核紀錄及相關文件紀錄，以讓其使用結果可以接受持續稽查。

6. 第三方供應商監察

本集團對第三方供應商採取有效的管控措施，確保第三方供應商滿足本集團關於供應商管理的要求及本管理辦法內的風險管理原則。

7. 道德操守及公平性

本集團遵守《銀行營運守則》、《公平待客約章》、普及金融的精神、秉持客戶為本的原則及其他適用的監管規定和消費者保障原則，以及本集團的企業價值及道德標準。人工智能應用不應涉及違反集團企業價值及道德標準、會明顯損害客戶作出知情決定的能力或其基本權利、對客戶構成嚴重損害、涉及剝削弱勢社群。

本集團亦釐定模型演算法的關鍵原則，以緩解歧視及對待客戶不公平或有偏差等風險，確保人工智能應用能提供客觀、一致、秉持道德操守及公平的結果，並遵守有關歧視的法例。

8. 透明度與披露

本集團確保人工智能應用的使用者於使用人工智能應用或服務前已向其作出妥善、準確及易於理解的披露，確保傳達的資訊清晰及簡單易明，提供適當的透明度，包括但不限於披露服務是以人工智能技術推動並解釋人工智能的用途與目的、相關風險及限制。

9. 定期重檢及持續監察

基於人工智能模型存在出現「模型漂移」或「模型衰退」的風險，本集團制訂適用及與風險相應的監測工具，並為人工智能應用的使用者提供反饋途徑。透過數據及 / 或使用者反饋進行持續監察，並適時分析結果及作出相應跟進。除對模型表現的持續監察，亦會對模型表現是否滿足既定標準以及其合規性進行定期重檢。

10. 數據保護

本集團按照集團內有關資訊安全的政策及管理辦法等規定處理，以確保遵守《個人資料（私隱）條例》（《私隱條例》）、私隱專員公署就《私隱條例》提供實務性指引而發布的相關實務守則，及其他適用的監管規定。

11. 網絡安全

為應對潛在對本集團人工智能應用的惡意網絡攻擊，本集團配置相應網絡安保防禦措施，並密切留意新興的惡意網絡攻擊手法及適時重檢有關安保防禦機制，及進行相應提升防禦工作。

12. 風險緩解和應變計劃

本集團對人工智能的活動進行適當的風險控制（例如「人在環中」），並就人工智能應用制定應變計劃，或在必要時暫停人工智能應用並觸發回退流程。