

(翻譯本)

本局檔號： B1/15C  
B9/67C

致： 所有認可機構  
行政總裁

敬啟者：

## 應用生成式人工智能的消費者保障

謹就認可機構在面向客戶的服務中應用生成式人工智能(GenAI)提供一套有關消費者保障的指導原則。

因應大數據分析及人工智能(BDAI)的發展，香港金融管理局（金管局）於2019年11月5日「認可機構就應用大數據分析及人工智能的消費者保障」通告中發出一套指導原則（《2019年BDAI指導原則》），涵蓋四大範疇，分別是管治與問責、公平性、透明度與披露，以及資料私隱與保障（參閱附件1摘要<sup>1</sup>）。從金管局近期所作統計調查反映BDAI的應用案例蓬勃增長可見（參閱附件2調查結果摘要），有關指導原則對銀行及客戶帶來裨益，並促進BDAI在整體銀行業界的良好發展。更重要的是，《2019年BDAI指導原則》進一步增強了客戶對應用BDAI的銀行服務的信心。

近月來，金管局注意到銀行業對在其業務操作中應用GenAI的興趣漸增。GenAI作為BDAI的一種形式，它能夠以大量數據為基礎而生成新的內容，例如文字、圖像、音訊、視訊、編碼或其他媒體。目前，銀行業界在應用GenAI方面仍處於早期階段，現階段的應用大多集中於提升銀行運作效率，例如內部聊天機械人及編碼。然而，GenAI具備創作內容的能力，預示著它可能在面向客戶的活動中被銀行業更廣泛地應用。潛在應用範圍包括但不限於客戶聊天機械人、定製化產品及服務的開發與交付、針對性的銷售及市場推廣，以及用於財富管理與保險的機械理財顧問。GenAI在面向客戶活動中的應用，將對消費者保障產生影響。

作為BDAI的一個分支領域，GenAI基本上具備與BDAI相近的風險元素。因此，就面向客戶的應用程式的消費者保障而言，金管局預期所有認可機

---

<sup>1</sup> 有關認可機構應用人工智能(包括 GenAI)的風險管理指引，請參閱金管局2019年11月1日「應用人工智能的高層次原則」通告；有關指引會因應市場發展及實際經驗不時作出更新。

構會將《2019年BDAI指導原則》延伸落實於GenAI的應用，並繼續秉持與所涉風險相稱的風險為本方法。此外，由於GenAI採用複雜模型，潛在風險例如缺乏可解釋性及「幻覺」（即生成看似現實但實際上不正確、不完整、缺乏重要訊息或與情境無關的結果）可能會導致對客戶更重大的影響。因此，金管局就四大範疇各自制定下述新增原則，旨在確保在採用GenAI於面向客戶的應用程式時，適當的保障措施已到位，以保障消費者權益。

## 1. 管治與問責

認可機構董事局及高級管理層須對所有以GenAI推動的決定及程序負責，並在認可機構的管治、監察及問責架構下，通過適當的委員會，全面考慮GenAI應用程式對客戶的潛在影響。當中包括確保：

- (a) 清楚界定面對客戶的GenAI應用程式的涵蓋範圍，使GenAI不會被應用於非擬定環節；
- (b) 就以負責任方式使用面向客戶的GenAI應用程式，制定妥善政策與程序以及相關管控措施；及
- (c) 在使用面向客戶的GenAI應用程式的早期階段，認可機構應適當核實GenAI模型，特別是應採用「人在環中」方式，即在決策過程中保留人類控制權，確保模型生成的結果準確可靠，不會有所偏頗。

## 2. 公平性

認可機構應確保GenAI模型能為客戶提供客觀、一致、秉持道德操守及公平的結果，當中包括確保：

- (a) 模型產生的結果不會對任何客戶或客戶群組造成不公平的偏見或不利影響。認可機構應考慮在GenAI模型中可採用的不同方法，例如把某些類別的數據匿名化、運用能全面及持平地代表相關總體的數據集、在核實及檢視過程中作出調整以消除偏見（例如採用「人在環中」方式）等；及
- (b) 在使用面向客戶的GenAI應用程式的早期階段，認可機構應向客戶提供退出使用GenAI的選項(opt-out)，並在客戶的要求下，盡可能對GenAI產生的決定作出人手干預。若基於某些原因無法提供退出使用GenAI的選項，認可機構應提供相關渠道，讓客戶要求檢視GenAI產生的決定。隨著科技的不斷發展，能夠得出與退出使用GenAI的選項

相同效果的其他措施亦會被接受。

### 3. 透明度與披露

認可機構應透過向客戶作出妥善、準確及易於理解的披露，就GenAI應用程式提供予客戶適當的透明度。因此，認可機構應向客戶披露其GenAI的應用，當中包括向客戶說明應用GenAI模型的用途與目的，以及有關模型的限制，促進客戶對模型產生的結果的了解。

### 4. 資料私隱與保障

認可機構應實施有效的客戶資料保障措施。若GenAI應用程式涉及收集及處理個人資料，認可機構應遵守《個人資料（私隱）條例》及參考個人資料私隱專員公署就GenAI發出的相關建議及良好行事方式，包括分別於2021年8月18日及2024年6月11日發出的《開發及使用人工智能道德標準指引》與《人工智能(AI)：個人資料保障模範框架》。

#### 積極使用BDAI及GenAI以加強消費者保障

BDAI，尤其是GenAI，具備潛力優化產品特點及將客戶細化區分至個人層面，讓銀行能以高效率及定製化的方式，精準地為特定客戶設計及推廣特定產品，創造業務機遇。基於同樣的邏輯，金管局鼓勵認可機構探討使用BDAI，包括GenAI，來加強消費者保障。一些例子可能包括：識別屬弱勢社群並需要較多保障與教育的客戶；識別可能需要更多資訊或解說以幫助其更清楚了解產品特點、風險和條款及細則的客戶；或向進行涉及較高潛在風險交易的客戶發出警示以防範詐騙等。

若對本通告有任何查詢，請聯絡葉卓詠女士(2597-0495)或梁文傑先生(2878-1186)。

助理總裁（銀行操守）  
區毓麟

2024年8月19日

副本送： 香港銀行公會主席  
存款公司公會主席  
財經事務及庫務局局長（收件人：杜奕霆先生）

**金融管理局於 2019 年 11 月 5 日發出的通告**  
**「認可機構就應用大數據分析及人工智能(BDAI)的消費者保障」**  
**所述的指導原則**

## 1. 管治與問責

認可機構董事局及高級管理層須對所有以 BDAI 作出的決定及程序負責，因此，當中應包括確保：

- (a) 制定適當的管治、監察及問責架構，並以文件記錄；
- (b) BDAI 模型（包括任何演算法）須具適當程度的可解釋性（即不能以「黑盒」作藉口），而該認可機構能了解有關模型；
- (c) 遵守《銀行營運守則》、《公平待客約章》及其他適用的監管規定的消費者保障原則，如同提供傳統銀行產品及服務時一樣。BDAI 應用程式亦應與該認可機構的企業價值和道德操守準則一致，當中包括秉持客戶為本的文化及原則；及
- (d) BDAI 應用程式須經適當的核實才投入運作，其後亦須作出持續檢視，確保模型本身、所用數據及結果可靠、公平、準確及相關。

## 2. 公平性

認可機構應確保BDAI模型能為客戶提供客觀、一致、秉持道德操守及公平的結果，當中包括確保：

- (a) 遵守適用的法律，包括有關歧視的法例；
- (b) 不會無理地拒絕客戶使用基本銀行服務，否則有違普及金融的精神；
- (c) 顧及客戶的財務能力、狀況及需要，包括其數碼認知水平；
- (d) 利用 BDAI 作出決定的相關模型應穩健可靠，並已適當顧及所有相關變數；及

- (e) 按需要容許人手干預以避免不負責任的放貸決定（例如涉及較高風險或較大影響的自動化決定的個案）。

### 3. 透明度與披露

認可機構應透過向客戶作出妥善、準確及易於理解的披露，就BDAI應用程式提供予客戶適當的透明度，當中應包括：

- (a) 在提供服務前向客戶清楚表明該服務已使用BDAI技術，並解釋相關風險；
- (b) 向客戶作出適當披露，讓他們了解認可機構使用客戶資料的方式；
- (c) 為客戶設立機制，以供查詢及要求檢視BDAI應用程式所作的決定，並確保有關使用BDAI的產品及服務的投訴處理及補償機制是方便及公平；
- (d) 按客戶要求及若情況適合，向客戶解釋所使用的是哪些類別的數據，以及哪些因素會影響或有關模式會如何影響BDAI作出的決定。惟此要求不須應用於監察及防範欺詐、洗錢或恐怖分子資金籌集活動的系統；
- (e) 推行適當的消費者教育，增進他們對有關應用BDAI技術的銀行服務的認識；及
- (f) 確保向客戶傳達的資訊清晰及簡單易明。

### 4. 資料私隱與保障

認可機構應實施有效的客戶資料保障措施，當中應包括：

- (a) 若 BDAI 應用程式涉及收集及處理個人資料：
  - 確保遵守《個人資料（私隱）條例》（《私隱條例》），包括6項保障資料原則、個人資料私隱專員（私隱專員）就遵守《私隱條例》提供實務性指引而發布或批准的相關實務守則，以及本地與海外其他適用的法定或監管要求；

- 參考私隱專員發出 BDAI 及金融科技相關的良好行事方式，包括「道德問責框架」(Ethical Accountability Framework，英文版)、為配合該框架的「數據管理問責、數據影響評估和監督模式」，以及「金融科技資料單張」；及
- (b) 從最初已考慮把資料保障融入產品或系統的設計內（即「貫徹私隱的設計」），並且只會收集最少量資料及儲存最短時間（即「資料最少化」）；及
- (c) 若須要求獲得同意就使用 BDAI 技術的銀行產品或服務收集及使用個人資料，應確保有關同意盡量清晰及容易明白，並確保客戶在充分知情下作出此同意。

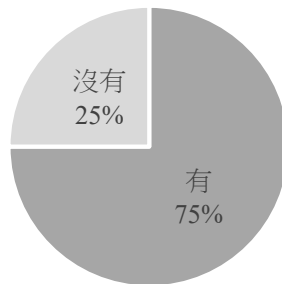
## 就應用大數據分析及人工智能的消費者保障 統計調查

於 2024 年 5 月，金管局就認可機構應用大數據分析及人工智能(BDAI)，當中包括生成式人工智能(GenAI)，進行統計調查。是項調查涵蓋 28 間主要服務零售客戶的認可機構，所得主要觀察摘要如下：

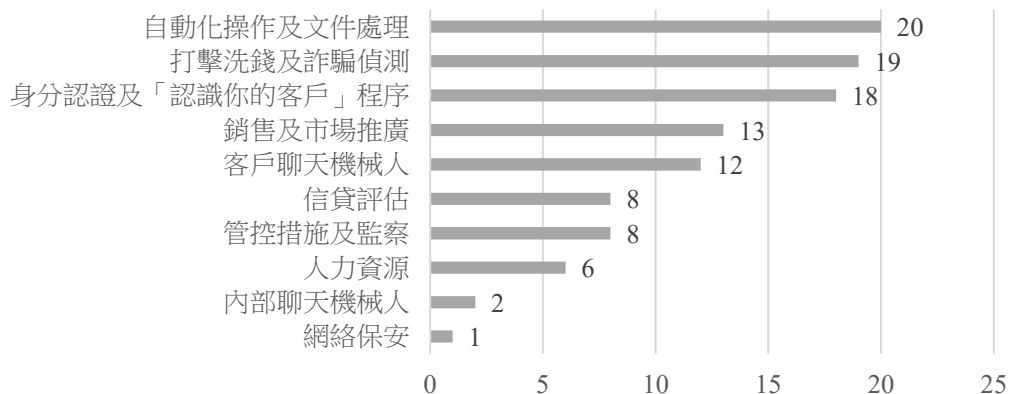
### A. BDAI 的應用

75%的受訪認可機構表示已經或計劃應用 BDAI 以提供一般銀行產品和服務，以及進行日常運作。實際案例涵蓋面向客戶的活動（身分認證、客戶聊天機械人、信貸評估）、中間部門職能（打擊洗錢及詐騙偵測、管控措施及監察），以及後勤支援部門職能（自動化操作及文件處理）。各類第三方服務提供者廣泛參與提供 BDAI 應用程式。

#### A1. 已經或計劃應用 BDAI 的認可機構數目



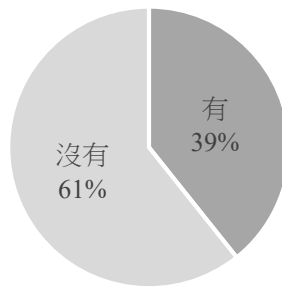
#### A2. BDAI 實際案例分類



## B. GenAI 的應用

39%的受訪認可機構表示已經或計劃應用 GenAI 以提供一般銀行產品和服務，以及進行日常運作。大多數實際案例用於內部業務職能，例如資料撮要及翻譯、編碼及內部聊天機械人。在面對客戶的活動應用 GenAI 技術仍處於早期階段，有關案例較集中於草擬客戶回覆，以及協助擬備銷售及市場推廣資料。大多數被採用的 GenAI 程式均源自都是來自第三方服務提供者的現成方案。

### B1. 已經或計劃應用 GenAI 的認可機構數目



### B2. GenAI 實際案例分類

