AI and Ethics: Ensuring the Responsible Use of Generative AI in Banking

人工智能與道德：確保銀行業負責任地使用生成式人工智能

With generative AI (genAI) adoption becoming widespread, and opportunities to use genAI in the banking and finance sector continuing to proliferate, this article seeks to deliberate on its ethical and responsible use.

隨著生成式人工智能(genAI)被廣泛應用，銀行和金融業使用genAI的機會愈來愈多，本文特此探討有道德和負責任地使用人工智能。

In November 2022, OpenAI debuted ChatGPT. Since then, numerous chatbots powered by genAI have emerged – such as Baidu's ERNIE Bot, Anthropic's Claude and Google's Bard – allowing the general public to interact with AI in a conversational manner. Despite a recent dip in traffic, genAI chatbots remain popular. In August 2023 alone, ChatGPT drew over 1.4 billion visits worldwide.

The banking sector stands to benefit from genAI's transformative power. A study by consultancy firm McKinsey estimates that full adoption of the technology's potential use could deliver up to USD340 billion of value, or 4.7% of the banking industry's annual revenue if the use cases McKinsey used to form the estimate were fully implemented. Research by another consultancy firm, Accenture, suggests that 54% of the current tasks in the banking sector have potential for transformation using genAI. Despite its potential, however, genAI poses privacy and ethical risks. Furthermore, if left unchecked, these risks could mutate into serious issues.

GenAI gaining ground in the banking industry

First, what is genAI? According to an answer provided by ChatGPT, it is "a subset of artificial intelligence that uses machine learning techniques to generate new data or content, such as images, text or music that is similar to or based on the training data it has been fed".

Banks have long deployed chatbots to handle straightforward customer queries. As mobile platforms flourished in the 2010s, AI chatbots evolved to offer direct assistance to users or to guide them to relevant help resources.

GenAI, however, is revolutionising this domain. The appeal of genAI is multifaceted: it can deliver personalised customer support, provide contextual marketing ideas, ensure round-the-clock services, and help staff better understand customers' needs. In fact, an investment bank has just lodged a patent application for a genAI service for equity selection.

In the banking and finance environment, genAI covers the following areas:

1. **Risk management and compliance**: GenAI has proven to be valuable in risk management. GenAI models can enhance anti-money laundering surveillance, spot patterns of suspicious transactions, raise alerts promptly and minimise missing cases. An example is a US-based bank that incorporated genAI into its fraudulent and suspicious activity detection systems, resulting in a marked decrease

自OpenAI於2022年11月推出新研發的聊天機械人ChatGPT後，眾多由genAI驅動的聊天機械人應運而生，例如百度的文心一言（ERNIE Bot）、Anthropic的Claude和Google的Bard，讓公眾能以對話的方式與人工智能互動。儘管最近使用率有所下滑，但genAI聊天機械人仍然備受歡迎。僅在2023年8月，ChatGPT的全球訪問量就超過14億次。

銀行業勢將從genAI的變革力量中受益。麥肯錫顧問公司的一項研究估計，genAI的潛在用途若得以全面發揮，可帶來高達3,400億美元的價值，即銀行業年度收入的4.7%。另一家顧問公司埃森哲的研究顯示，銀行業目前54%的工作都有可能利用genAI進行轉型。然而，儘管genAI潛力鉅大，它也帶來私隱和道德風險。這些風險如果不加以控制，可能會演變成嚴重問題。

銀行業採用GenAI日益普遍

首先，甚麼是genAI？根據ChatGPT提供的答案，它是「人工智能的一種分類，使用機器學習技術來產生新的資料或內容，如根據所接收的訓練數據，產生類近的圖像、文字或音樂」。

銀行界長期以來一直使用聊天機械人來處理簡單的客戶查詢。隨著流動平台在2010年代蓬勃發展，人工智能聊天機械人亦不斷演變，為用戶提供直接的協助或引導他們獲取相關有用的資源。

然而，genAI正在這個領域進行革命性的改變。genAI有多方面的吸引力：它可以提供個人化的客戶支援、提供配合不同情境的市場推廣意念、確保全天候服務，並協助員工更好地了解客戶的需求。事實上，一家投資銀行剛剛提出一項用於選擇股票的genAI服務專利申請。

就銀行和金融業而言，genAI涵蓋以下範疇：

1. **風險管理和合規**：GenAI在風險管理方面已被證實有重要價值。GenAI模型可以加強反洗黑錢監控、識別可疑交易模式、及時發出警告，並減少漏報情況。其中一個例子是一家美國銀行將genAI納入其欺詐和可疑活動檢測系統，從而大幅減少系統誤報並提高欺詐識別率。此外，genAI還可以簡化可疑交易報告的制訂，以及製作

in false positives and improved fraud detection rates. Moreover, genAI can simplify the compilation of suspicious transaction reports and strengthen staff training by simulating realistic cases. In compliance, genAI can help decipher dense regulatory texts, compare regulations across jurisdictions, and highlight areas for improvement.

2. **Business enhancement**: From a business standpoint, genAI helps banks in predicting economic trends and simulating practices at a macro level. At a micro level, it may influence every step of the lending process, from borrower analysis to market research and regular reviews.

3. **Operational improvements**: Operational improvements can be achieved in various business areas, including administration, human resources, technology, procurement and legal. Although these advancements are not exclusive to banking, the sector's innovative nature often paves the way for early adoption. A prime example of the power of genAI is content writing, which enables the generation of pitch books with insights drawn from a vast array of resources, resulting in significant time and resource savings.

Privacy pitfalls

While banks traditionally enjoy a high degree of trust from their customers in handling personal data, the introduction of genAI presents new privacy and ethical concerns.

> *"In addition to the Guidance, developers or users of AI systems are also recommended to adopt a personal data Privacy Management Programme (PMP), a management framework for the responsible collection, holding, processing, and use of personal data by an organisation."*

仿真個案來加強員工培訓。在合規方面，genAI可以幫助解讀繁瑣的監管文本、比較不同司法管轄區的法規，並突顯需要改進的地方。

2. **提升業務**：從商業角度來看，genAI有助銀行預測經濟趨勢，並模擬未來宏觀環境。在微觀層面上，它可影響貸款過程中的每一步驟，由借款者分析、市場研究以及定期審查。

3. **改善營運**：可以在各個業務範疇中改善營運，包括行政、人力資源、科技、採購和法律。儘管這些進展並非銀行業獨有，但銀行業追求創新的本質往往有助從業員及早採用新技術。例如，genAI擅於編寫，它能夠根據大量資料生成意念，從而編訂營運計劃書，有助節省大量時間和資源。

私隱陷阱

雖然銀行在處理個人資料方面一直獲得客戶高度信任，但genAI帶來了新的私隱和道德問題。

To dissect these issues, it is possible to map them against the Data Protection Principles (DPPs) in the Personal Data (Privacy) Ordinance that cover the entire lifecycle of personal data from collection, holding, processing, use to deletion.

Data Collection

The first privacy risk relates to data collection. GenAI-powered chatbots use "deep learning technology", which involves analysing massive volumes of unstructured data without supervision.

Whether banks build their own generative models or fine-tune existing models with their specific data, the process requires a vast amount of data, including personal data, which might originate from transaction records, customer profiles, financial statements and more. As the original purpose of collecting such personal data may not have included AI model training, this type of use potentially circumvents the DPPs governing collection and transparency (DPPs 1 and 5), which require personal data to be collected in a fair manner and on an informed basis.

要剖析這些問題，我們需要留意《個人資料（私隱）條例》中的保障資料原則，該等原則涵蓋個人資料收集、持有、處理、使用到刪除個人資料的整個生命周期。

資料收集

第一個私隱風險與資料收集有關。genAI支援的聊天機械人使用「深度學習技術」，往往在缺乏監察的情況下分析大量原始資料或數據。

無論銀行是自行建立生成式模型，或利用獨有的數據微調現有模型，過程都需要大量數據，包括個人資料。這些數據可能來自交易記錄、客戶資料、財務報表等。由於收集這類個人資料的原本目的可能並不包括人工智能模型訓練，因此有關的使用可能會違反了規管收集資料和資料透明度的保障資料原則（保障資料原則第1和第5原則），這些原則規定應當以公平的方式和在知情的基礎上收集個人資料。

使用限制

第二個私隱風險涉及與genAI工具的互動。

使用者輸入於genAI工具的資料可能包含姓名、身分證號碼和帳號等敏感個人資料，而這些資料有機會被用作模型的訓練數據，因而超出原先收集資料擬使用的目的。例如，虛擬助理可以將使用者輸入的資料與其他對話的資料進行比較，以便了解客戶的需求。但是，這可能構成濫用資料，違反了保障資料第3原則，該

"除參考《指引》外，我們亦建議人工智能系統的開發者或使用者訂定個人資料私隱管理系統，在此管治框架下負責任地收集、持有、處理和使用個人資料。"

> *"While banks traditionally enjoy a high degree of trust from their customers in handling personal data, the introduction of genAI presents new privacy and ethical concerns."*

Use limitation
The second privacy risk pertains to interactions with genAI tools.

User inputs, potentially encompassing sensitive personal data such as names, identity card numbers and account numbers, might be used beyond their original purpose and used as training data for the models. For instance, a virtual assistant could compare users' inputs with data from other conversations to better understand customers' needs. Such potential misuse might contravene the use limitation principle (DPP3), which stipulates that personal data should not be used for a new purpose without the prescribed consent of the data subject.

Furthermore, there is a risk of sensitive information leakage in the models' outputs. With the new training data mentioned above, the system might fine-tune its responses, inadvertently leaking personally identifiable information.

Access to and correction of data
The third privacy risk relates to the challenges over the rights of data subjects to access and correct their personal data (DPP6) and the retention of the data (DPP2). Given the sheer volume of training data, whether it is practicable for users to access or correct their data remains an issue.

Data security
Another privacy risk concerns data security. This arises from the storage of numerous user conversations in the systems. Furthermore, genAI systems might be attacked by "jailbreaking" prompts, which are specifically designed to bypass safeguards for malicious ends and might cause operational issues or data leakage. This would violate the data security principle (DPP4), which requires personal data to be protected against unauthorised or accidental access, processing, erasure, loss or use.

原則規定未經資料當事人同意，不得將個人資料用於新目的。

此外，模型在輸出結果時亦存在敏感資訊外洩風險。使用上述新訓練數據後，系統可能會微調給其他使用者的回應，從而無意中洩露可識別個人身份的資料。

資料的存取和更正
第三個私隱風險所帶來的挑戰，涉及資料當事人存取和更正自己的個人資料的權利（保障資料第6原則）和資料的保留（保障資料原則第2原則）。鑑於genAI訓練資料數量龐大，使用者存取或更正其資料是否可行，仍是一個問題。

資料安全
由於系統中儲存了大量用戶對話，數據安全是genAI的另一個私隱風險。此外，genAI系統可能會受到「越獄」攻擊（即指專為繞過防範惡意攻擊而設計的指令），導致操作出現問題或資料外洩，這將違反要求保護個人資料免遭未經授權或意外查閱、處理、刪除、喪失或使用的保障資料第4原則。

Broader ethical implications

Beyond privacy risks, genAI also poses broader ethical risks, including the following.

1. **Explainability**: While decisions made by traditional AI are often hard to interpret, this challenge is intensified in genAI. These models are trained on vast and diverse datasets, making it extremely difficult to trace certain outputs to specific data inputs, let alone to explain the technical process to lay users.

2. **Accuracy**: GenAI can produce confident but incorrect statements. In banking, this might entail a misestimated loan risk, or erroneous outputs to regulators. Another cause of genAI's inaccuracy is its underperformance in numerical analysis compared with text. This is a challenge for banks, where numbers are of the essence.

3. **Risk amplification**: A note published by the International Monetary Fund suggests that the use of genAI could escalate systemic risks. This stems from the "herd mentality" that arises from different banks using the same systems, and a potential inclination towards high-risk suggestions from the models due to a lack of robust risk management during training.

更廣泛的道德問題

除了私隱風險，genAI還帶來更廣泛的道德風險，包括：

1. **可解釋性**：GenAI比起傳統人工智能所作的決策更加難以解釋。GenAI這類人工智能模型以龐大且多樣化的資料集進行訓練，極難從輸出的資料追溯到特定的輸入資料，更不用說向非專業用戶解釋當中的技術流程。

2. **準確性**：GenAI能產生有說服力但錯誤的句子。在銀行業，這可能會導致錯誤估計貸款風險，或向監管機構提供錯誤資料。GenAI不準確的另一個原因，是數字分析的表現較文字分析差。這對建基於數字的銀行業來說，是個挑戰。

3. **風險擴大**：國際貨幣基金組織發表的報告指出，genAI的使用可能會加劇系統性風險。這是由於不同銀行採用相同系統，或引致「羊群心理」。另外，由於genAI訓練期間缺乏健全的風險管理，銀行可能會傾向於採納模型中的高風險建議。

Navigating the changes: Existing guidance and emerging AI regulations

Responding to the risks presented by the development and use of AI, authorities around the world have proposed or have already put in place regulations and laws.

In Mainland China, the "Interim Measures for the Management of the Services by Generative AI", the world's first regulations specific to AI-generated content, became effective in August 2023. The measures set out the obligations of genAI service providers to ensure, among others, the quality of training data and to adopt measures to prevent minors from becoming addicted to AI-generated content.

Elsewhere, the EU is planning to regulate AI, including genAI, with an Artificial Intelligence Act, which, if enacted, will introduce a risk-based approach and a new regulator. In Canada, the AI and Data Act is being considered, while the UK launched a consultation on its pro-innovation regulatory AI approach in March 2023.

Despite the variation in approaches, the common goal of these jurisdictions is to create an environment in which AI can evolve and operate in a manner that respects privacy, protects data, curtails bias and champions transparency.

Apart from regulations, governments and regulators have issued guidances and recommendations on the development and deployment of AI.

In August 2021, the Office of the Privacy Commissioner for Personal Data, Hong Kong (PCPD) published the "Guidance on the Ethical Development and Use of Artificial Intelligence" (Guidance). The Guidance outlines frameworks for deploying AI in a privacy-friendly manner that minimises ethical risks, while striking a balance between fostering innovation and ensuring the protection of personal data.

To this end, the PCPD proposed three sets of recommendations: three data stewardship values, seven ethical principles and a four-step practice guide.

The data stewardship values – being respectful, beneficial and fair – underscore the importance of treating individuals as human beings, not as mere data sets. AI should work for the benefit of the broader community while minimising any possible harm. Fairness should

應對變化：現行指引和新人工智能法規

為應對人工智能發展和使用帶來的風險，世界各地有關當局都正在提議或已經制定了一些法規和法律。

在中國內地，全球首個專門針對人工智能生成內容的法規《生成式人工智能服務管理暫行辦法》於2023年8月實施。該《辦法》訂定了人工智能服務提供者的責任，包括確保訓練數據的質量，並採取有效措施，防範未成年人用戶沉迷生成式人工智能提供的內容。

在其他地區，歐盟正計劃通過《人工智能法案》，監管包括genAI的人工智能。法案如獲實施，將引入基於風險為本的措施，並成立新的監管機構。另外，加拿大正考慮訂定《人工智能及數據法案》，而英國則在2023年3月展開支持創新的人工智能監管方案諮詢。

儘管方法各異，但各個司法管轄區的共同目標，都是創造一個讓人工智能以尊重私隱、保障資料、減少偏見和具透明度的方式發展和運作的環境。

> *"雖然銀行在處理個人資料方面一直獲得客戶高度信任，但genAI帶來了新的私隱和道德問題。"*

permeate throughout the process of using the technology and the results it generates. Any differential treatments should be justified.

The seven ethical principles of accountability; human oversight; transparency and interpretability; data privacy; fairness; beneficial AI; and reliability, robustness and security, align with internationally recognised principles in this area.

The four-step practice guide recommends various safeguards on the basis of four business processes, namely, the establishment of an internal governance structure, conducting comprehensive risk assessments, execution of AI model development and system management, and fostering of communication with stakeholders, including the organisation's employees and customers.

In addition to the Guidance, developers or users of AI systems are also recommended to adopt a personal data Privacy Management Programme (PMP), a management framework for the responsible collection, holding, processing and use of personal data by an organisation.

除法規外，政府和監管機構亦有就人工智能的開發和使用發布指引和提出建議。

香港個人資料私隱專員公署便於2021年8月發出《開發及使用人工智能道德標準指引》，概述以保護私隱為原則的使用人工智能框架，以減少道德風險，並在促進創新和保障個人資料之間取得平衡。

為此，私隱專員公署提出三套建議：三項數據管理價值、七項道德原則和分為四個步驟的實務指引。

數據管理價值，即尊重、互惠和公平，強調應視個體為人而不僅是數據集的組成部分。人工智能應該對社會有裨益，同時盡量減少任何可能的傷害。科技的使用過程及其產生的結果都應公允，任何有差別的待遇都應該有充份的理據支持。

七項道德原則與國際公認的原則一致，包括問責、人為監督、透明度和可解釋性、數據私隱、公平、有益的人工智能、以及可靠、穩健和安全。

實務指引提出四個步驟，就四項業務流程提出各種保障措施，包括建立內部管治架構、進行全面風險評估、實行人工智能模型的開發和人工智能系統的管理，以及促進與持份者（包括機構的員工和客戶）的溝通。

除參考《指引》外，我們亦建議人工智能系統的開發者或使用者訂定個人資料「私隱管理系統」，在此管治框架下負責任地收集、持有、處理和使用個人資料。

The PMP comprises three components. The first component is organisational commitment, which hinges on top-management buy-in, the appointment of a dedicated data protection officer, and the creation of a reporting mechanism directly to senior management.

The second component is programme controls, which encapsulate control measures such as the formation of a personal data inventory, the establishment of data handling policies, and the application of risk assessment tools, among others.

The third component is ongoing assessment and revision of the PMP. Organisations should devise a plan for oversight and review, and periodically revise their programme controls.

The three components collectively enhance data security for any organisation, allow effective mitigation of the privacy risks associated with the use of AI, ensure compliance with the requirements of the Personal Data (Privacy) Ordinance, and build trust with stakeholders.

Joining hands in ensuring the ethical and responsible use of AI

More than 10 years ago, the emergence of smartphones gave rise to internet banking apps, which were initially met with customer hesitation owing to fears of security risks such as identity theft and insecure data storage. However, as measures, including robust regulations, have been introduced to better safeguard personal data, customer acceptance has increased and mobile banking has become increasingly common.

Today, as genAI advances, the banking industry and society at large are again confronted with a crucial decision. Should we shun genAI just because it poses potential risks, thus stripping customers of a chance to enjoy better services? Or should we deploy the technology in a manner that ensures its ethical and responsible use, thereby harnessing the technology's benefits without compromising our valued principles and rights? The answer is clear. Together, let us collaboratively craft a proper regulatory framework and establish norms to enable the development and use of AI in a privacy-friendly and ethical manner. BT

「私隱管理系統」由三個部分組成。第一個組成部分是機構的決心，這取決於高級管理層的支持、委任專責的保障資料人員，以及建立直接向高級管理層負責的匯報機制。

第二個組成部分是系統管控措施，包括建立個人資料庫存、訂定處理個人資料的內部政策，以及採取風險評估工具等。

第三部分是對管理系統的持續評估和修訂。機構應制定監督和檢討計劃，並定期修訂其系統管控措施。

這三個組成部分合力增強任何機構的資料保安，有效緩解使用人工智能相關的私隱風險，確保符合《個人資料（隱私）條例》的要求，並與持份者建立信任。

攜手確保道德與負責任地使用人工智能

十多年前，智能手機的出現催生了網路銀行應用程式。由於擔心保安風險，如身份被盜用和資料儲存不安全，客戶最初對這些應用程式心存疑慮。然而，隨著當局推出嚴格監管及其他相關措施，個人資料保障更趨完善，客戶接受程度不斷提高，手機銀行也愈來愈普遍。

隨著genAI不斷發展，現時銀行業和整個社會再次面對一個關鍵性的決定。是否因為有潛在風險而迴避genAI，從而剝奪客戶享受更好服務的機會？抑或恪守道德並負責任地使用人工智能技術，從而在不損害我們重視的原則和權利的基礎上，善用人工智能技術的好處？答案顯而易見。讓我們同心協力，制定適當的監管架構，建立規範，以尊重私隱和合乎道德的方式開發和使用人工智能。BT

ABOUT THE AUTHOR
作　者　簡　介

Ada CHUNG Lai-ling
Barrister
Privacy Commissioner for Personal Data

鍾麗玲大律師
個人資料私隱專員