De-risking Generative AI for Banking and Finance
銀行金融業界如何解除 Gen AI風險

Artificial intelligence (AI) technologies have penetrated all industries and are now being used for myriad applications. In financial services, many organisations are looking to build AI-driven platforms to establish a competitive advantage, uplift productivity and enhance customer engagement. Today, AI is seen as one of the most disruptive technologies impacting businesses. But there are gaps to fill.

人工智能科技已滲透各行各業，並在無數個不同的場景中應用。在金融服務業，許多機構都希望透過構建人工智能驅動的平台，建立競爭優勢、提升生產力、加強與客戶聯繫。人工智能被視為今天對商業活動影響最大的科技之一，但仍然存在一些有待改善之處。

As AI adoption and use have grown, so too has awareness of the various risks and challenges of deploying AI — from unintended bias to determining accountability for outcomes. Most financial services adopters agree that AI risks and the uncertainty in managing compliance are slowing AI adoption and progress.

There appears to be limited implementation of specific actions to help mitigate those risks, even by the most skilled adopters. Some of the key questions on the minds of leaders and managers are:

- How to differentiate Generative AI from AI?
- How do we demonstrate de-risking of AI across all use cases?
- How can we embed ethics into AI builds and use?
- How do we leverage existing governance structures to demonstrate compliance?
- How do we balance the enablement and governance of AI?

Cutting-edge AI competencies

In its earliest form, AI systems were designed to mimic human intelligence to perform tasks such as decision-making, pattern recognition and problem-solving. These data-driven systems are often used by financial institutions to streamline processes and optimise performance across different business functions. For instance, in banking, AI-powered systems are deployed to detect fraud, analyse transaction patterns to identify suspicious activities, and prevent financial losses. In investment banking, AI algorithms are used for high-frequency trading, leveraging real-time market data to make split-second trading decisions, and capitalise on market opportunities. In insurance, AI is employed for risk assessment and underwriting, and analysing vast datasets to determine insurance premiums and policy eligibility criteria.

人工智能的應用日趨普及，人們也日益認識使用人工智能所涉的風險和機遇，包括無意中產生的偏見，以及如何決定生成結果的負責方等。大部分採用人工智能的金融服務機構都同意，人工智能風險和管理合規事宜的不確定性，導致機構放緩對人工智能的採納和應用進程。

即使是最擅長應用人工智能的機構，似乎也缺乏有助減少相關風險的具體行動。企業領導和管理人員的主要疑問包括：

- 如何區分人工智能和生成式人工智能？
- 如何證明所有應用場景的人工智能風險都已解除？
- 如何在構建和使用人工智能的過程中融入倫理道德價值？
- 如何運用現有管治架構證明合規？
- 如何平衡人工智能的賦能作用和管治？

人工智能的高超本領

早期的人工智能系統，作用是模仿人類從事決策、辨認模式和解決難題等工作。金融機構往往使用這些以數據為本的系統精簡程序，提升各業務部門的表現。例如銀行界利用人工智能驅動的系統偵測詐騙活動、分析交易模式以辨識可疑交易，以及避免財務損失；投資銀行利用人工智能演算法進行高頻交易，參考實時市場數據，以極高速度作買賣決定，捕捉市場機會；保險業則運用人工智能作風險評估及風險核保，並且分析大量數據，以決定保單保費及投保資格標準。

> “These stakeholders play a crucial role in defining the organisation’s AI governance framework, ensuring the establishment of appropriate risk management measures, roles and responsibilities.”

The emergence of Generative AI (GenAI) marks a departure from these conventional applications, emphasising the creation of new, synthetic content that closely resembles human-generated data. At the heart of GenAI lies a diverse array of generative models, each tailored to produce specific types of content such as text, images and music or video. Notable among these models are Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs) and Transformer-based models like GPT (Generative Pre-trained Transformer). Unlike traditional AI, which has limited interpretability on data conceptual representations, GenAI can decode semantic associations from contextual backgrounds and generate new content by a similar pattern.

Large language models (LLMs) have emerged as powerful tools capable of processing and generating text at scale. These models have revolutionised natural language processing tasks, enabling unprecedented levels of linguistic understanding and generation. By training on vast amounts of text data, LLMs have demonstrated remarkable proficiency in generating coherent and contextually relevant text, further blurring the lines between human and machine-generated content.

有別於這些傳統的應用場景，生成式人工智能的應用重點，在於創造與人類產生的數據非常相似的新合成內容。生成式人工智能由多種不同的生成模型構成，每個模型專門產生特定種類的內容，例如文字、圖像、音樂或影片。較值得注意的模型有生成對抗網路、變分自編碼器，以及變換模型如生成式預訓練變換模型(GPT)等。傳統人工智能對數據概念的分析能力有限，生成式人工智能則可從語境中解讀語意關聯，然後以類似模式生成新內容。

大型語言模型的誕生，提供了有力的工具，協助處理和生成大量文字。這些模型大大改變了自然語言處理的工作，把理解和生成語言的能力提升至前所未有的程度。以大量文本數據訓練後，大型語言模型表現卓越，可生成通暢而有條理的文字，進一步縮小了人類與機器產生的內容之間的差別。

Impacts on financial institutions

As financial institutions explore the potential of GenAI, LLMs have already demonstrated various applications in banking and finance, including sentiment analysis of customer feedback to gauge market trends and customer satisfaction, automated generation of financial reports and summaries for enhanced decision-making efficiency, and natural language understanding for customer service chatbots to provide personalised assistance and streamline communication. Recent developments on multi-modality large models will further enable financial institutions to tackle complex tasks requiring comprehensive understanding.

Another notable application of GenAI is Stable Diffusion, a generative model used for creating high-quality images. Stable Diffusion can generate high quality synthetic images for fraud detection systems, enabling the creation of realistic synthetic data for training AI models to enhance privacy and security measures while preserving clients' data integrity.

對金融機構的影響

在金融機構探討生成式人工智能的潛力之際，大型語言模型在銀行金融業已有多種應用，包括分析客戶回應的情緒，以判斷市場趨勢和客戶滿意度；自動製作財務報告及摘要，以提升決策效率；以及協助客戶服務聊天機械人理解自然語言，以便為客戶提供個人化的協助，使溝通更精簡。多模態大型模型的近期發展，將可將一步協助金融機構處理需要融會貫通的複雜工作。

生成式人工智能的另一種值得注意的應用，是Stable Diffusion擴散模型，這是用以創作優質影像的生成式模型。Stable Diffusion可以為詐騙活動偵測系統生成優質合成影像，以便創作逼真的合成數據，用以訓練人工智能模型，從而加強保障私隱與安全，同時維持客戶數據的完整性。

Governance towards responsible AI

As financial institutions increasingly integrate AI and GenAI into their operations however, the need for robust governance frameworks becomes paramount. The ethical and regulatory considerations surrounding AI adoption necessitate comprehensive governance structures that span the entire lifecycle of AI applications, from design and development to deployment and ongoing monitoring.

To address these issues, financial institutions must establish clear accountability and oversight mechanisms, starting from the board and senior management levels. These stakeholders play a crucial role in defining the organisation's AI governance framework, ensuring the establishment of appropriate risk management measures, roles and responsibilities.

From the initial design and development phases to ongoing monitoring and review, ethical considerations must be integrated seamlessly into every aspect of AI implementation. This fosters a strong ethical culture within the organisation, translating ethical principles into practical guidelines and ensuring continuous training and awareness among relevant stakeholders.

人工智能管治

不過，隨着越來越多金融機構在業務運作中應用人工智能和生成式人工智能，建立健全的管治架構便成為首要任務。人工智能的運用，涉及倫理道德和法規方面的考慮，機構必須設立完善的管治架構，涵蓋人工智能應用的整個生命周期，由設計與開發，以至落地實施和持續監察。

為處理這些事宜，金融機構必須設立清晰的問責和監督機制，由董事局和高級管理層做起。這些持份者在設定機構的人工智能管治架構方面起着關鍵作用，負責制訂恰當的風險管理措施，界定各相關人員的角色與職責。

由最初的設計與開發階段，到持續監察與檢討的整個過程，人工智能應用的每個方面均應結合倫理道德考慮。這樣便可在機構內建立道德色彩濃厚的文化，把倫理道德原則化為實際指引，確保為相關持份者持續提供培訓，增強其意識。

Ensuring trustworthy AI

Financial institutions can use the following trust domains in Deloitte's Trustworthy AI™ framework to explore the types of risks they might face when deploying Generative AI:

- **Fairness and impartiality**: Limiting bias in AI outputs is a priority for all models, whether machine learning or generative. The root in all cases is the latest bias in the training and testing of data. Inconsistent outcomes and performance in applications could diminish end user trust in the tool and in the business itself.
- **Transparent and explainable**: Given the capacity for some Generative AI models to convincingly masquerade as human, there is often a need to explicitly inform the end users that they are conversing with a machine. More broadly, to trust the model and its outputs, stakeholders and end users need an understanding of how input data is used, an opportunity to opt-out, obscure or restrict that data, and an accessible explanation of automated decisions and how they impact the user.
- **Safe and secure**: Generative AI can be susceptible to harmful manipulation such as divulging confidential

確保人工智能可以信任

德勤設計的可信任人工智能Trustworthy AI™框架中列明六大關鍵信任域，可幫助金融機構了解應用生成式人工智能時機構自身可能遇到的風險種類：

- **公平公正**：使用各種類型的人工智能模型，不管是機器學習還是生成式人工智能，最重要的是減少輸出成品中的偏見。問題的根源，是培訓和數據測試過程中的偏見。應用時輸出成品不一致，表現不穩定，會削弱使用者對工具本身甚至對機構的信任。
- **透明並可解釋**：有些生成式人工智能模型模仿人類時十分逼真，因此往往有需要明確告知使用者，與他們對話的是個機器。一般來說，持份者和使用者希望了解輸入的數據會作什麼用途，可以選擇離開、遮蔽或限制使用該數據，也希望獲得有關自動化決策的解釋，了解這種決策對使用者的影響，才會對人工智能模型及其輸出成品產生信任。

> “這些持份者在設定機構的人工智能管治架構方面起著關鍵作用，負責制訂恰當的風險管理措施，界定各相關人員的角色與職責。”

information and creating misinformation. To promote GenAI safety and security, businesses need to weigh and address factors around cybersecurity and the careful alignment of GenAI outputs with business and user interests.

- **Accountable**: Although AI accountability is squarely a human domain, Generative AI makes accountability much more complicated. Regardless of whether the enterprise uses an in-house or vendor model, there must be a clear link between the Generative AI model and the business deploying it.
- **Responsible**: For all the good it can be used to promote, GenAI use cases can also lead to harms and disruption. What is judged to be responsible deployment by one organisation may not be judged the same by another. Enterprise leaders must determine for themselves whether a GenAI use case is a responsible decision for their organisation.
- **Privacy**: Data used to train and test Generative AI models can contain sensitive or personally identifiable information that needs to be obscured and protected. As with other types of AI, the organisation needs to develop cohesive processes for managing the privacy of all stakeholders; they can remove personal data, use synthetic data or even prevent end users from inputting personal data into the system.

- **安全可靠**：生成式人工智能可能被不當利用，例如洩露機密資料、產生虛假資料等。為保持人工智能安全可靠，機構須權衡並處理網絡保安事宜，並致力確保生成式人工智能的輸出成品與機構和使用者的利益一致。
- **問責**：人工智能問責毫無疑問屬於人類的範疇，但生成式人工智能使問責問題更複雜。機構使用的不管是內部研發的模型，還是供應商提供的模型，生成式人工智能模型與使用模型的機構必須有明確聯繫。
- **負責任的使用**：使用生成式人工智能有許多好處，但也有機會帶來害處和干擾。某機構認為是負責任的使用，另一機構可能有不同看法。企業領導必須自行評定，對其機構而言，運用生成式人工智能是否負責任的決定。
- **私隱**：用以培訓和測試生成式人工智能模型的數據，可能含有敏感或個人可識別資訊，須予遮蔽及保護。正如使用其他類型的人工智能一樣，機構須訂立嚴謹的程序，保障所有持份者的私隱；他們可以移除個人資料、使用合成的數據，甚至阻止最終使用者把個人資料輸入系統。

Benchmark governance structure

Moreover, financial institutions need to leverage existing governance structures to effectively manage the challenges posed by AI. By mapping existing risk management frameworks and committee structures to the specific risks and requirements associated with AI, organisations can ensure cohesive oversight and alignment with regulatory mandates. This involves active engagement from key stakeholders, including board members, senior management, data scientists, ethicists and legal and compliance personnel.

Furthermore, embedding ethics into the development and deployment of AI systems requires a multi-faceted approach that spans the entire AI lifecycle. It begins with establishing a strong ethical culture and governance framework at the highest levels of the organisation. Boards and senior management must define and uphold ethical principles and values that guide the development and use of AI systems, aligning them with relevant laws, regulations and industry standards.

Translating ethical principles into actionable guidelines and processes is essential for ensuring effective implementation. This involves providing comprehensive training to relevant

管治架構標準

此外，金融機構須善用現有的管治架構，來有效應對人工智慧帶來的挑戰。將現有的風險管理框架和委員會架構，與人工智能相關的特定風險與要求相互對照，機構便可確保進行統一的監督，符合法規的要求。這過程需要董事局成員、高級管理層、數據科學家、倫理學家及法律合規人員等主要持份者的積極參與。

再者，在開發和利用人工智能系統時結合倫理道德，須採取多方面的做法，涵蓋人工智能的整個生命周期。首先是在機構最高層建立牢固的倫理道德文化和管治架構。董事局和高級管理層必須清楚界定並維護指導人工智能系統開發和使用的倫理道德原則和價值，使其符合相關法律、法規和行業標準。

要確保有效實施管治架構，機構必須把倫理道德原則轉化為可實踐的指引和程序。這包括為數據科學家、開發人員、業務使用者和風險經理等相關團隊提供全面的培訓，說明其工作的倫理道德意義，以及如何在實際工作上應用倫

> *“By mapping existing risk management frameworks and committee structures to the specific risks and requirements associated with AI, organisations can ensure cohesive oversight and alignment with regulatory mandates.”*

teams, including data scientists, developers, business users and risk managers, on the ethical implications of their work and how to apply ethical guidelines in practice. For example, data scientists must be equipped to detect and mitigate biases in data and models, while business users should understand the potential impact of AI decisions on customers.

Once AI systems are deployed, ongoing monitoring and review are essential to ensure continued adherence to ethical principles. Effective communication and redress mechanisms should be in place to address any adverse impacts of AI decisions on customers or other stakeholders.

Defence strategy

The “Three Lines of Defence” model can be applied to govern AI risks effectively.

- The first line of defence, comprising business units and AI developers, is responsible for identifying, assessing and mitigating risks in line with approved policies and standards.
- The second line, consisting of risk management and compliance functions, provides oversight and challenge to ensure adherence to ethical principles and regulatory requirements.
- The third line, internal audit, provides independent assurance on the effectiveness of the AI governance framework.

> *“將現有的風險管理框架和委員會架構，與人工智能相關的特定風險與要求相互對照，機構便可確保進行統一的監督，符合法規的要求。”*

理道德指引。例如數據科學家必須具備有關技能，在數據和模型中偵測和減少偏見；而業務使用者則應了解人工智能決策對客戶可能產生的影響。

一旦使用人工智能系統，機構必須持續監察和檢討，確保能一直符合倫理道德原則。機構應設立有效的溝通和申訴機制，處理人工智能決策對客戶或其他持份者造成的不良影響。

防線策略

機構可運用「三道防線模型」，有效地管理人工智能風險。

- 第一道防線由業務單位和人工智能開發者組成，負責按照核准的政策及標準去識別、評估和緩減風險。

By integrating AI governance into existing risk management structures and processes, financial institutions are able to demonstrate a cohesive and consistent approach to managing AI risks across the organisation. This enables them to leverage existing expertise, resources and oversight mechanisms while ensuring ethical considerations and regulatory requirements for AI are adequately addressed.

Conclusion

Balancing enablement and governance in AI adoption is of utmost importance for financial institutions, requiring a nuanced approach that integrates innovation with ethical conduct and regulatory compliance. By adopting a risk-based approach to governance, engaging stakeholders effectively and streamlining processes, organisations can facilitate responsible AI adoption, driving sustainable growth and competitive advantage in an AI-driven landscape. This strategy, which embeds ethics into development and leverages existing governance structures, ensures that financial institutions are able to navigate the complexities of AI and GenAI adoption while fostering trust and positioning themselves for long-term success. BT

- 第二道防線包括風險管理及合規職能，負責監察和提出質疑，確保符合倫理道德原則和法規要求。
- 第三道防線是內部審核，就人工智能管治架構的效能提供獨立保證。

金融機構在現有的風險管理架構和程序中結合人工智能管治，便可證明在整個機構採用嚴謹一致的方法管理人工智能風險。這樣一來，機構便可善用現有專門知識、資源和監督機制，同時確保與人工智能相關的倫理道德考慮和法規要求得到適當處理。

結語

在使用人工智能的過程中平衡賦能作用和管治，對金融機構來說極其重要，必須謹慎行事，將創新與倫理道德行為和監管合規性結合起來。藉着在管治上採取風險為本的做法，有效地與持份者聯繫，精簡流程，機構便可使人工智能得到負責任的使用，在人工智能驅動的環境下，促進可持續發展，加強比較優勢。這個策略將倫理道德融入發展，並善用現有的管治架構，可確保金融機構能夠應對採用人工智能及生成式人工智能的複雜問題，同時獲得信任，達至長遠成功。BT

ABOUT THE AUTHORS 作者簡介

Reginia CHAN
Banking & Capital Markets
Partner
Deloitte China

陳穎思
德勤中國銀行業及資本市場
合夥人

Dr Leo MA
Dean of Deloitte AI Institute
(Hong Kong)

馬培煒博士
德勤人工智能研究院(香港)院長

Kenneth YU
Partner
AI and Data Specialist
Deloitte China

余伯钧
德勤中國人工智能及數據業務
合夥人