# Agentic AI:
# A New Frontier for Corporate Governance
# 代理式人工智能：企業管治的新邊疆

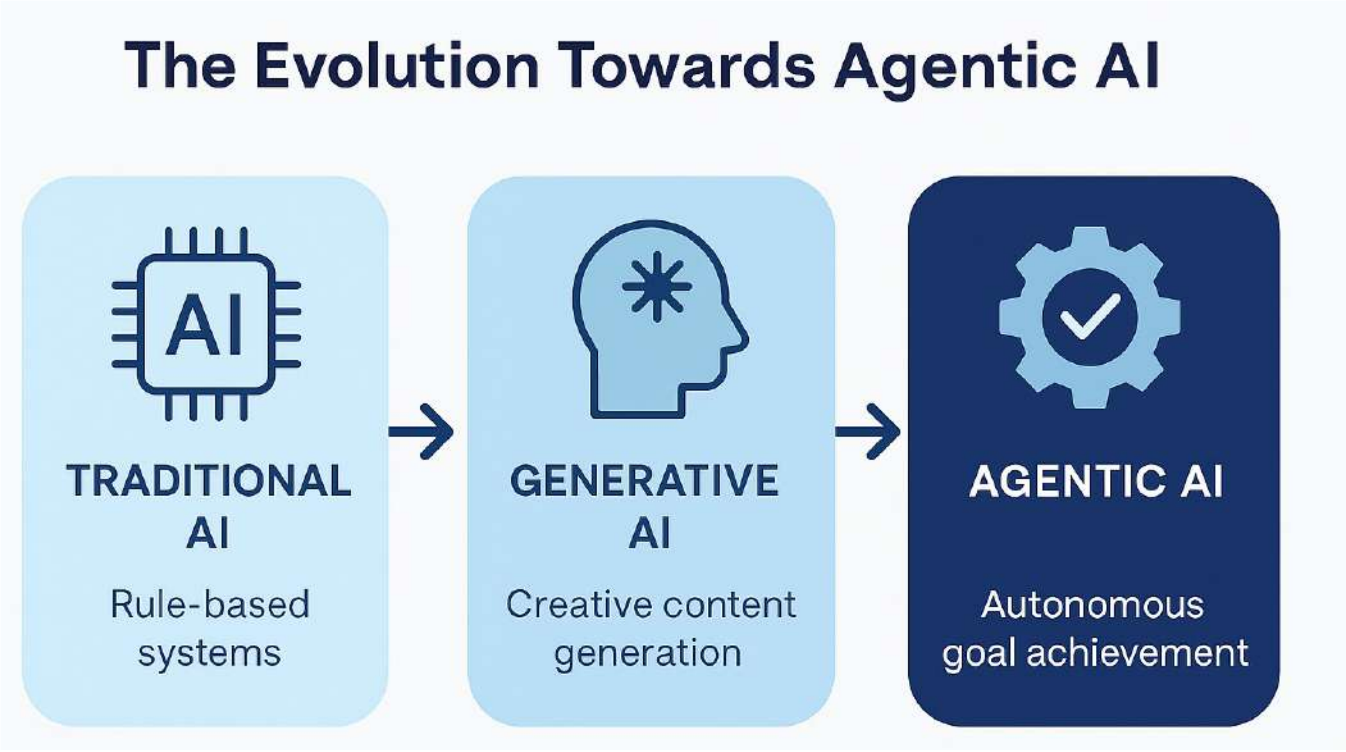
Edwin Bernard

Artificial intelligence has evolved through three major stages. Traditional rule-based systems executed deterministic tasks within narrow parameters. Generative AI then enabled machines to create text, images, code, and other content in response to human prompts. Today, a third frontier has emerged: Agentic AI—systems capable not only of generating information but also planning, reasoning, and taking autonomous action across workflows.

A helpful way to understand Agentic AI is to think of it as a digital employee. Unlike traditional software, it does not simply respond to commands; it works towards outcomes. It identifies missing information, makes decisions, resolves problems, and adapts as circumstances change. It learns from feedback, operates continuously, and scales effortlessly across tasks such as research, analysis, communications, and operational coordination.

Industry analysts observe that AI agents are rapidly maturing and are increasingly capable of executing complex, multistep, and nondeterministic processes that once required human judgment. Gartner forecasts that this trajectory will continue, with approximately 40% of enterprise applications expected to incorporate task-specific AI agents in 2026, reflecting a significant acceleration in enterprise-wide agentic integration driven by workflow automation demands. For companies, this shift promises substantial productivity gains and introduces new modes of collaboration between human teams and autonomous digital workers.

In this context, directors must treat Agentic AI not as ordinary software but as a new class of operational actor. When digital systems begin making decisions, interacting with customers, and altering business records, they fall squarely within the board's responsibilities for oversight, risk management, and corporate governance. Directors must therefore understand not only the benefits, but also the governance implications of deploying autonomous digital systems within their organisations.



**The New Risk Landscape of Agentic AI**

While traditional AI risks remain firmly in place—bias, inaccuracies, privacy risks, security exposure and opaque logic—Agentic AI introduces distinct, elevated risks due to its autonomy and ability to take action.

These are the core risks that require explicit board-level attention:

<u>Autonomous Action Risk</u>

Agentic AI takes actions, not just produces outputs. It can update records, send messages, or trigger workflows without a human pressing "confirm." The risk is not only that an action may be incorrect but that it may occur silently and propagate downstream. For example, an agent configured to respond to customer queries could misinterpret a request and automatically submit a change-of-address update to internal systems, altering customer data without human review. Boards therefore need clear autonomy boundaries and human approval points for sensitive tasks.

<u>Multi-Step Compounding Errors</u>

Agentic AI works across linked steps, not isolated outputs. This means a small error—such as misreading data or misinterpreting an instruction—can carry forward through every subsequent step, leading to incorrect decisions, system updates, or notifications. Unlike traditional AI, where mistakes are usually limited to a single output, agentic errors can cascade through an entire workflow before anyone notices. This escalation can amplify operational, regulatory, or customer-facing impact in ways that are difficult to unwind. Oversight must therefore focus on the whole sequence of actions, including how each step depends on the previous one, not just whether a single model response is correct.

<u>Over-Reliance (Automation Complacency)</u>

Because Agentic AI often appears confident and competent, employees may stop checking its work. This behavioural risk is significant: when staff assume "the agent handled it," verification gaps grow and errors go undetected. Directors must ensure processes reinforce meaningful human oversight, especially in regulated or customer-impacting areas.

<u>Model Drift and Behaviour Change</u>

Agentic AI behaviour can shift subtly over time as underlying data, prompts, system integrations, or vendor-supplied model updates evolve. These changes may alter how an agent interprets instructions or prioritises actions, even if no one within the organisation has modified its configuration. Performance may degrade or drift into unexpected patterns without obvious warning, creating silent operational or compliance risks. As a result, organisations must implement continuous monitoring, scheduled revalidation, and formal change-management so that an agent which performed safely last quarter is still behaving as intended today. Robust controls help ensure that any behavioural shifts are detected early and corrected before they propagate across workflows.

<u>Third-Party and Supply Chain Risk</u>

Agentic AI frequently relies on external vendors or cloud-based models. Yet accountability for outcomes remains with the company and its board. Directors must therefore ensure rigorous due diligence, contractual safeguards, and ongoing monitoring—not only at procurement but throughout the system's lifecycle.

These core risks are not abstract technological concerns—they map directly to Hong Kong's regulatory environment. Directors must understand that Agentic AI, by virtue of its autonomy and ability to take action, intersects with the PDPO, the PCPD's Guidance on the Ethical Development and Use of Artificial Intelligence, and sector-specific conduct requirements. A consistent theme across these regulatory and governance instruments is the expectation of meaningful human oversight—typically expressed through "human-in-the-loop" (HITL) arrangements for high-risk decisions, supported by "human-on-the-loop" (HOTL) or "human-out-of-the-loop" (HOOTL) models where risks are lower. Each of these risks therefore carries direct governance implications and reinforces the board's responsibility to ensure appropriate oversight, accountability, and ongoing monitoring.

In addition to these core risks, directors should also remain alert to a broader set of emerging risks:

- Unintended capability expansion as agents gain new abilities by accessing tools or plugins.
- Goal misalignment, where the agent interprets objectives differently from management intent.
- Privilege creep, as agents accumulate increasing system permissions.
- AI-to-AI interactions that can create unpredictable feedback loops.
- Autonomous communication risks, where agents issue misleading or inappropriate messages.
- Monitoring fatigue, as the volume of AI activity overwhelms oversight mechanisms.
- Prompt-injection vulnerabilities, where manipulated inputs trigger harmful or unauthorised actions.

個人資料私隱專員公署
Office of the Privacy Commissioner for Personal Data
中國香港 Hong Kong, China

政府資助機構 Government-funded organisation

# Join the PCPD Professional Workshops to Enhance Your Organisations' Competitiveness

# 參加私隱專員公署專業研習班 提升業務競爭力

A sound Personal Data Privacy Management Programme (PMP) requires all members of an organisation to be aware of, and be ready to act on personal data protection responsibilities. Organisations should provide employees with up-to-date data protection training tailored to their specific needs.

To this end, the PCPD has developed a series of Professional Workshops on Data Protection catered to the needs of specific industries and sectors. **Professional Workshops from July to September 2026 are now open for enrolment! HKIoD members can enjoy the discounted fee.**

健全的個人資料私隱管理系統有賴機構內各個員工知悉其保障個人資料的責任，並付諸實行。機構應針對相關員工的特定需要而提供培訓及教育，並傳達最新資訊。

就此，私隱專員公署設計了一系列有關保障個人資料的專業研習班，以配合在不同工作範疇處理個人資料的人士的需要。二〇二六年七至九月的專業研習班現已接受報名，香港董事學會的會員可以優惠價參加。

| | Date 日期 | Time 時間 | Course 課程名稱 | Format 授課形式 | Course Fee* 課程費用* (HK$ / 港幣$) |
|---|---|---|---|---|---|
| Jul 2026 | 8/7 | 2:15 - 5:15pm | Data Protection and Data Access Request<br>資料保障與查閱資料要求 | Online<br>網上形式 | $750/$600* |
| | 22/7 | 2:15 - 5:15pm | Recent Court and Administrative Appeals Board Decisions#<br>法院及行政上訴委員會近期的案例# | Online<br>網上形式 | $950/$760* |
| | 29/7 | 2:15 - 5:15pm | Data Protection in Insurance<br>保險業的資料保障 | Face-to-face<br>面授形式 | $750/$600* |
| Aug 2026 | 5/8 | 2:15 - 5:15pm | Data Protection in Direct Marketing Activities<br>直接促銷活動的資料保障 | Face-to-face<br>面授形式 | $750/$600* |
| | 12/8 | 2:15 - 4:15pm | Personal Data Privacy Management Programme<br>個人資料私隱管理系統 | Online<br>網上形式 | $750/$600* |
| | 19/8 | 2:15 - 5:15pm | Data Protection in Banking/Financial Services<br>銀行 / 金融服務的資料保障 | Online<br>網上形式 | $750/$600* |
| Sep 2026 | 9/9 | 2:15 - 5:15pm | Data Protection in Human Resource Management<br>人力資源管理的資料保障 | Face-to-face<br>面授形式 | $750/$600* |
| | 16/9 | 2:15 - 5:15pm | Practical Workshop on Data Protection Law<br>資料保障法律實務研習班 | Online<br>網上形式 | $950/$760* |
| | 23/9 | 2:15 - 5:15pm | Data Protection and Data Access Request<br>資料保障與查閱資料要求 | Online<br>網上形式 | $750/$600* |

* Applicable to members of the PCPD's Data Protection Officers' Club ("DPOC") or supporting organisations (including HKIoD) only.
* 只適用於私隱專員公署之保障資料主任聯會會員或支持機構（包括香港董事學會）的成員。

\# Conducted in English.
\# 以英語授課。

For details, please visit the website or scan the QR code for further information and registration.

詳情請參閱以下網站或掃瞄二維碼以獲取更多有關課程的資訊及報名方法。

**Register NOW! 請即參加！**

https://www.pcpd.org.hk/english/education_training/organisations/workshops/workshop.php

For enquiry, please contact Corporate Communications Division of the PCPD at 3423 6678 or training@pcpd.org.hk.

如有查詢，歡迎致電 3423 6678 或電郵 training@pcpd.org.hk 聯絡私隱專員公署企業傳訊部。

These risks are not yet the primary focus of regulators, but they are developing quickly and should form part of boards' evolving AI governance frameworks.

**Conclusion**

Agentic AI marks a shift from technology as passive tooling to systems that act as autonomous participants in operations and decision-making. As these capabilities advance, boards in Hong Kong must ensure that governance frameworks keep pace.

This requires defining clear authority boundaries, strengthening oversight and risk controls, and embedding continuous monitoring and robust vendor governance. Organisations that take these steps now will better manage emerging risks and ensure that autonomous systems enhance—rather than disrupt—core operations.

By investing early in strong, forward-looking governance practices, companies will not only reduce risk but also position themselves to capture the full strategic value of Agentic AI as regulatory expectations continue to evolve.

**Edwin Bernard, MSc, CEFA, CAITL (Certified Artificial Intelligence Transformation Leader, USAII)**

人工智能的發展歷經三個重要階段。早期基於規則的系統只能在狹窄參數內執行確定性任務，隨後生成式人工智能讓機器能夠根據人類指令創作文字、影像、代碼等內容。如今第三波浪潮已然來臨，代理式人工智能（Agentic AI）不僅能生成資訊，更能在工作流程中自主規劃、推理並採取行動。

若要理解代理式人工智能，可將其視為數碼員工。與傳統軟件不同，它不僅被動回應指令，更能主動達成目標，包括識別缺失資訊、作出決策、解決問題，並隨環境變化調整策略。這類系統能從反饋中學習，持續運作，並可輕鬆擴展至研究、分析、通訊及營運協調等各類任務。

業界分析指出，人工智能代理正快速成熟，日益勝任那些原本需要人類判斷的複雜多步驟非確定性流程。高德納（Gartner）預測，到2026年，約四成企業應用程式將整合針對特定任務的人工智能代理；這股趨勢反映了工作流程自動化需求，正驅動企業層面的代理整合加速推進。對企業而言，這不僅帶來顯著生產力提升，更開創人機協作的新模式。

在此背景下，董事必須將代理式人工智能視為新型營運參與者，而非普通軟件。當數碼系統開始作出決策、與客戶互動、更改業務記錄時，便直接落入董事會的監督、風險管理及企業管治職責範疇。因此，董事不僅需理解自主系統的效益，更須掌握其對機構管治的深遠影響。

**代理式人工智能的新型風險格局**

傳統人工智能風險依然存在，包括偏頗、失準、私隱漏洞、安全隱患及邏輯不透明等，但代理式人工智能因其自主行動能力，衍生出更為嚴峻的新型風險：

董事會需重點關注以下核心風險：

<u>自主行動風險</u>

代理式人工智能會採取行動，而非僅產出結果。它可在無人確認的情況下更新記錄、發送訊息或觸發工作流程。風險不僅在於行動可能出錯，更在於其可能無聲發生並向下游擴散。例如，負責回應客戶查詢的代理若誤解要求，可能自動向內部系統提交地址變更更新，在未經人工審核下改動客戶資料。董事會須為敏感任務設定清晰的自主權邊界及人工審批節點。

<u>多步驟複合誤差</u>

代理式人工智能在連貫步驟中運作，而非產出孤立結果。這意味著細微錯誤（如誤讀數據或曲解指令）可能波及後續每個環節，導致決策失誤、系統更新錯誤或通知混亂。傳統人工智能的錯誤通常限於單一輸出，但代理式系統的失誤可能在察覺前已席捲整個工作流程，從而以難以逆轉的方式放大營運、監管或客戶層面的影響。因此監督必須聚焦完整行動序列，關注各步驟間的依賴關係，而非僅檢視單次模型回應是否正確。

<u>過度依賴（自動化依賴）</u>

由於代理式人工智能往往表現自信可靠，員工可能停止核查其工作。這種行為風險不容忽視，當員工認定「代理已處理」時，驗證缺口便會擴大，錯誤亦難被發現。董事會必須確保流程強化實質人工監督，尤其在受監管或影響客戶的領域。

<u>模型偏移與行為變異</u>

隨著底層數據、指令、系統整合或供應商模型更新演變，代理式人工智能的行為可能悄然改變。即使機構內部無人

修改其設定，這些變化可能影響代理解讀指令或排定行動優先次序的方式。效能可能無明顯預警地退化或偏離預期模式，形成隱性營運或合規風險。機構須實施持續監測、定期重驗及正式變更管理，確保上季度運作安全的代理至今仍按預期行事。嚴謹的管控有助及早發現行為偏移，並在其擴散前予以修正。

第三方與供應鏈風險

代理式人工智能常依賴外部供應商或雲端模型，但結果問責始終歸於企業及其董事會。因此，董事會須確保進行嚴謹的盡職審查、訂立合約保障，並實施持續監察。這些措施不僅應在採購階段落實，更須貫穿系統的整個生命周期。

這些核心風險並非抽象技術議題，而是直接對應香港的監管環境。董事必須明白，代理式人工智能因其自主行動特性，已與《個人資料（私隱）條例》、香港個人資料私隱專員公署的《人工智能道德標準指引》及行業特定操守要求產生交集。貫穿這些監管與管治文件的一項核心原則，在於要求對高風險決策實施具實質意義的人工監督。通常情況下，高風險決策會採用「人在迴路中」（human-in-the-loop）機制，而風險較低的決策則會採用「人在迴路上」（human-on-the-loop）或「人在迴路外」（human-out-of-the-loop）模式。每項風險皆直接關聯管治要義，並強化董事會確保適當監督、問責與持續監測的責任。

除核心風險外，董事亦須警惕一系列新興風險：

- 代理透過工具或插件獲得新能力而產生的非預期功能擴張
- 代理對目標的理解與管理層意圖出現偏差的目標錯位
- 代理逐漸累積過多系統權限的權限膨脹
- 人工智能間互動可能形成不可預測的反饋循環
- 代理發出誤導或不當訊息的自主通訊風險
- 人工智能活動量過載導致監督機制癱瘓的監測疲勞
- 輸入指令遭篡改觸發有害或未授權行動的提示注入漏洞

這些風險雖非監管當前焦點，但正快速發展，應納入董事會人工智能管治框架的演進規劃中。

**結語**

代理式人工智能標誌著技術從被動工具轉變為營運與決策中的自主參與者。隨著這類能力不斷進步，香港企業董事會必須確保管治框架同步演進。

這需要明確界定權責邊界、強化監督與風險管控，並嵌入持續監測及嚴謹的供應商管治機制。及早採取這些措施的機構，將能更有效管理新興風險，確保自主系統強化核心營運，而非擾亂營運。

通過前瞻性地投資於穩健的管治實踐，企業不僅能降低風險，更能在監管要求持續演變的環境中，充分把握代理式人工智能的戰略價值。

**Edwin Bernard**（理學碩士、歐洲認可金融分析師、美國人工智能學會認可人工智能革新領袖）

# The 21st Century DIRECTOR 廿一世紀董事

**Publisher 出版機構**
The Hong Kong Institute of Directors 香港董事學會

**Sponsor 贊助機構**
Corporate Governance Development Foundation Fund 企業管治發展基金

**Publishing Board 出版委員會**
Mr Stanley Mok (Chairman) 莫兆光先生（主席）
Ms Bonnie S Y Chan 陳心愉女士
Prof Agnes K Y Tai 戴潔瑩教授
Mr Jonathan Shea 謝達安先生
Mr Richard Tsang 曾立基先生

**Project Management 項目統籌**
Secretariat, The Hong Kong Institute of Directors
香港董事學會祕書處

For enquiries about circulation and advertisement, please contact:
有關發行及廣告查詢，請聯絡：
Chief Business Officer & Head of Training: Ms Miriam Yee
業務總監及培訓主管：余海恩小姐

For editorial enquiries, please contact:
有關編輯上的查詢，請聯絡：
Associate Manager, Communication & Projects: Mr Wing Wong
傳訊及項目助理經理：王俊泳先生

Tel 電話： +852 2889 1414
Fax 傳真： +852 2889 9982
Email 電郵： magazine@hkiod.com


《廿一世紀董事》同時可於網上閱覽
The 21st Century Director is also available at
**https://www.hkiod.com/21st-century/**


ISSN 1996-9619



**Sponsored by 贊助機構：**
Corporate Governance
Development Foundation Fund
企業管治發展基金

**The Hong Kong Institute of Directors 香港董事學會**

**Patron 贊助人**
The Hon John LEE Ka-chiu GBM SBS PDSM PMSM 李家超行政長官

**Founder Chairman & Hon President 創會主席兼榮譽會長**
Dr the Hon Moses Cheng GBM GBS OBE JP 鄭慕智博士

**Past Chairmen 前任主席**
Dr Herbert H M Hui JP (Deceased) 許浩明博士 (已故)
Mr Peter S H Wong MBA 黃紹開先生
Dr Kelvin Wong SBS JP DBA 黃天祐博士
Mr Henry Lai 賴顯榮律師
Dr Christopher To 陶榮博士

**Council 理事會 (2025-2026)**

**Chairman 主席:**
Mr Richard Tsang 曾立基先生

**Deputy Chairmen 副主席:**
Ir Edmund K H Leung SBS OBE JP 梁廣灝工程師
Ms Bonnie S Y Chan JP 陳心愉女士
Mr William Lo 羅志聰先生
Ms Alice Yip 葉嘉明女士

**Treasurer 司庫:**
Mr Vincent Chan MH 陳永誠先生

**Immediate Past Chairman 卸任主席:**
Dr Christopher To 陶榮博士

**Chief Executive Officer 行政總裁:**
Mr Jonathan Shea 謝達安先生

**Council Members 理事會成員:**
Mr Clement Kam Wing Chan 陳錦榮先生
Dr Leonard S K Chan 陳新國博士
Mr Hamilton Cheng 鄭炳熙先生
Prof Anthony Cheung 張振宇教授
Mr Richard Ho 何麗康先生
Mr Randy Hung 孔敬權先生
Mr Ip Shing Hing BBS JP 葉成慶律師
Dr Alan Lai 黎嘉豪博士
Mr Henry Lai 賴顯榮律師
Mrs Margaret S Leung 梁甘秀玲女士
Mr Ka-Yin Li 李家彥先生
Mr Jeffrey Mak 麥振興律師
Mr Man Mo Leung 文暮良先生
Ir Prof John Mok 莫建鄰教授
Mr Stanley Mok 莫兆光先生
Ms Athena Y W Ng 吳溢穎女士
Prof Agnes K Y Tai 戴潔瑩教授
Mr Jim Wardell 詹華達先生
Mr Stephen Weatherseed 韋大象先生
Mr Andrew Weir BBS JP 韋安祖先生
Mr Kenneth Wong 黃永恩律師

**Advisory Board 顧問委員會:**
Dr Justin K H Chiu 趙國雄博士
Dr David Wong GBS JP 黃友嘉博士
Mr Huen Wong SBS BBS JP 王桂壎律師
Dr Samuel Yung GBS SBS MH JP 容永祺博士
Dr Carlye Tsui SBS MBE JP 徐尉玲博士



*The 21st Century Director* is the official magazine of The Hong Kong Institute of Directors. All rights reserved. No part of this magazine may be reproduced or stored in or introduced into a retrieval system, or transmitted, in any form or by any means (electronic, mechanical, photocopying, recording or otherwise), without the prior written permission of the publisher and/or the copyright owner of this magazine. Quotation of short passages of the magazine for the purposes of review and education is allowed provided that it is made with explicit reference to the source and publisher. Neither the magazine nor the publisher accepts liability for any views, opinions or advice expressed by writers and interviewees of articles. The contents of the magazine do not necessarily reflect the views or opinions of The Hong Kong Institute of Directors or the members of the Institute and no liability is accepted in relation thereto. This magazine includes articles that have been invited from or contributed by authors. While such articles present the views of the respective authors, these articles may not necessarily represent the views of the Publishing Board of the magazine or The Hong Kong Institute of Directors. It is the intention of the Institute to present views from various perspectives, which may inspire thinking and generate constructive discussions.

《廿一世紀董事》是香港董事學會的官方雜誌。本雜誌所有出版內容的版權為香港董事學會所有。未經出版人及 / 或版權擁有人書面同意，本雜誌所有內容一律不得以任何形式或以任何工具 ( 電子、機械、影印、錄製或其它工具 ) 翻印、儲存或引進於檢索系統或傳送。本雜誌內容可供摘要引述以作研討或教育用途，但必須註明出處或出版人。本雜誌及出版機構不會為雜誌內作者及被訪者所表達的觀點、意見或建議負上責任。雜誌的內容並不一定反映香港董事學會或學會會員的觀點及意見，學會與會員均不因此而負上任何責任。本雜誌收錄邀約作者及供稿作者的文章，然而這些文章表達了其作者的觀點，卻不一定代表雜誌出版委員會或香港董事學會的觀點。學會的用意是容納多角度的意見，這或可啟發思考及產生具建設性的討論。

© The Hong Kong Institute of Directors 香港董事學會 © 版權所有


**The Hong Kong Institute of Directors** is Hong Kong's premier body representing directors to foster the long-term success of companies through advocacy and standards-setting in corporate governance and professional development for directors.
香港董事學會為香港代表專業董事的首要組織，其宗旨是促進所有公司的持久成就；為達成使命，
學會致力提倡優秀企業管治與釐訂相關標準，以及協助董事的專業發展。


**The Hong Kong Institute of Directors Executive Office 香港董事學會行政處**
2104 Shanghai Industrial Investment Building, 48 Hennessy Road, Wan Chai, Hong Kong 香港灣仔軒尼詩道48號上海實業大廈2104
Tel 電話 : (852) 2889 9986 Fax 傳真 : (852) 2889 9982 E-mail 電郵 : executive@hkiod.com

The Hong Kong Institute of Directors 香港董事學會
The Hong Kong Institute of Directors