ISACA

人工智慧革命的機會與挑戰：風險管理之道

風險

目 錄

摘要

許多科幻電影的核心設定：當科技發展超越了人類智慧，並造成嚴重破壞且最終接管了人類。雖然這類電影還沒有在現實生活中上演，但最近ChatGPT的問世，感覺就像是開場白，而這也只是眾多生成式AI的工具之一。如今，蘋果共同創辦人史蒂夫．沃茲尼亞克（Steve Wozniak）和特斯拉執行長伊隆．馬斯克（Elon Musk）等知名科技領袖，都公開呼籲要求企業在業界進行風險評估前暫緩「大型人工智慧實驗」[1]。在他們的公開信中指出，如果沒有監督和智慧管理，大規模人工智慧專案「可能會對社會和人類構成嚴重的風險」，並要求暫時停止進一步的開發。這封信還強調了治理體系及專責人工智慧（AI）新監管機構的必要性。

儘管如此，也有些產業領袖認為，生成式AI的效益和風險都被誇大了[2]。實際上，我們面對的技術有限，這些技術既無法達到我們理想期待中的有用，也不至於如最遭想像中的那般強大。但有一點是肯定的：人工智慧（AI）已經席捲了我們的企業和社會，迫切需要資訊安全長（CISOs）、資訊風險經理、高階主管和資訊高階經理人跟上這快速發展的風險格局。

[1] Futureoflife.org，「暫停大型人工智慧實驗：一封公開信」，2023 年 3 月 22 日，https://futureoflife.org/open-letter/pause-giant-ai-experiments/

[2] Rundle, J.;" 網路安全主管應對人工智慧風險與潛在回報" 《華爾街日報》，2023 年 5 月 25 日，https://www.wsj.com/articles/cybersecurity-chiefs-navigate-ai-risks-and-potential-rewards-9138b76d

企業智慧與人工智慧的交會

就在幾個月前，人們還很難想像各行各業的專業人士會只用一個應用程式就能生成法律論證與建議、電腦程式碼、莎士比亞風格的十四行詩和臨床治療計畫。然而，數以百萬計的使用者已迅速適應了生成式AI工具所帶來的便捷新世界，例如OpenAI 的 ChatGPT[3] 和 Google 的 Bard。[4]儘管生成式AI取得了巨大的成功，用途廣泛，但大多數使用者並不了解生成式AI的風險。因此，隨著企業紛紛投入設計和利用人工智慧工具的白熱化競爭中，許多風險管理工作卻明顯得是落後了。

投入新興技術的建立和應用，不應以犧牲風險管理為代價。不幸的是，在大多數新技術應用中，這卻是常見的形況。還記得智慧型手機、3D列印或物聯網（IoT）等出現時的情形嗎？有些人可能會以為人工智慧（AI）的出現與這些技術性突破情形不相上下。

但實際上，它更像是汽車或網際網路的誕生：一股創新的火焰將席捲每個行業。

投入新興技術的建立和應用，不應以犧牲風險管理為代價。

人工智慧（AI）為我們與技術互動方式帶來了本質的轉變。以往過去那些為我們服務的傳統防護措施和安全協定，如今在一定程度上已經力不從心，因為生成式AI的普及和影響力只會越來越大。新的領域蘊含著新的風險，與以往的技術躍進不同，尋求權威指引的企業反而發現，如何深思熟慮地最大化人工智慧價值的同時，又巧妙地將風險降至最低，已成為關鍵的挑戰。

人工智慧的現況

儘管人工智慧（AI）能夠模仿人類的思維模式和語言，但它並不具備感知能力。大多數人們所稱的人工智慧（AI），其實是由機器學習技術或大型語言模型（LLMs）所組成。

例如，ChatGPT 在被餵入3000 億個詞彙[5]的資料集後，吸收了書籍、網站、維基百科以及其它來源，並透過「基於人類回饋的強化學習」（RLHF）進行訓練，隨著反饋不斷改善其回應的品質。Bard 也以類似的方式進行訓練，但也從網際網路上獲取資訊。

Megatron[6] 是微軟和輝達的合資企業，透過新穎的「平行化」技術進行訓練，並且有望憑藉著因能力的顯著提升而超越 ChatGPT。開發這些工具所使用的技術引發了一些問題，比如說關於LLMs中使用的資料來自哪裡，以及是否信任這些資料用於訓練人工智慧工具。

雖然像ChatGPT這樣的私人公司獨占了大部分媒體的關注，但新創公司如ChatSonic、Jasper、Wordtune等，正在幫助使用者將人工智慧（AI）視為一種全新的功能性科技，而不僅僅是單一工具。

[3] Openai.com，「ChatGPT 簡介」，2022 年 11 月 30 日， https://openai.com/blog/chatgpt

[4] bard.google.com，「試試 Bard，Google 的人工智能實驗」，https://bard.google.com/

[5] Iyer, A.，「ChatGPT 智慧的背後：3000 億單字、570 GB 資料」，《Analytics India 雜誌》，2022 年 12 月 15 日，https://analyticsindiamag.com/ behind-chatgpts-wisdom-300-bn-words-570-gb-data/

[6] Developer.nvidia.com，「Megatron-Turing 自然語言生成模型」，《NVIDIA Developer》，2022 年 10 月 3 日，https://developer.nvidia.com/mega-tron-turing-natural-language-generation

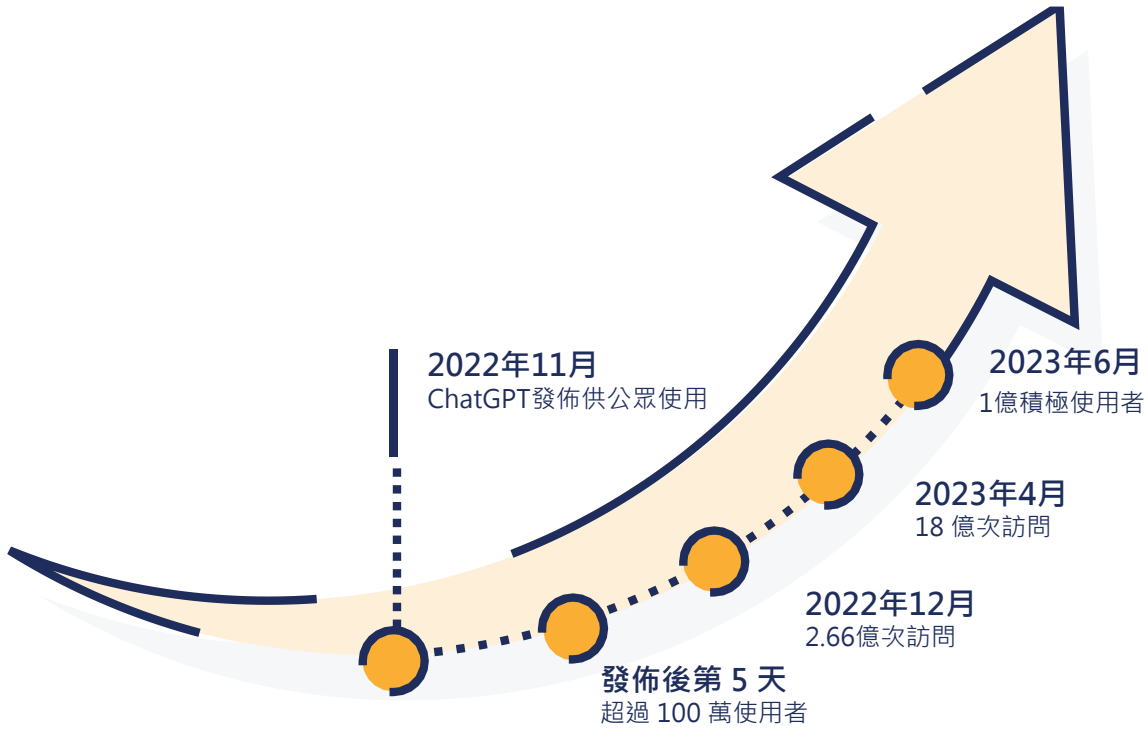
這部分是透過利用來自私人和公共部門的資訊貢獻，進行生成式AI預訓練來實現的。這些公司正在徹底改變我們與人工智慧（AI）的互動方式，以提供一系列工具和服務來滿足使用者的多樣化需求。為了回應使用者的反饋，開發者正在增加安全控制、文本分類與其他功能，以變得更具競爭力和實用性。

舉例來說，Bard 被設計成也可以作為個人助理，可以完成從預訂假期到制定餐點計畫等各種任務。ChatGPT 則被整合至 Microsoft Office 365 中作為 Microsoft 365 Copilot[7]。不久，任何使用 Microsoft Word、Excel 或 PowerPoint 的人都將能夠使用 Copilot 建立新的簡報、制定行銷策略，甚至將最複雜的財務數據濃縮成一份報告。

人工智慧（AI）的發展速度甚至讓非常有經驗的資深科技人員也眼花繚亂。幾乎每週都會出現對人工智慧（AI）的優勢、風險、局限性或潛力的新論述，但常常結論是自相矛盾。許多商業領袖選擇等待人工智慧（AI）塵埃落定後，再制訂正式的商業戰略，這不令人意外。儘管這似乎是最安全的途徑，但延遲本身也將會帶來風險。

出於這個原因，無論人工智慧革命以何種形式呈現，我們都知道人工智慧（AI）將持續發展，最安全、最明智的安全途徑是現在就開始適應。企業領袖應該假設人工智慧（AI）已經在企業內部使用，因為它的受歡迎程度在過去幾年中呈現指數性增長。圖 1 顯示了 ChatGPT 使用量爆炸性成長[8]的例子。

圖 1：AI 的使用突破音速



[7] Spataro, J.，「介紹 Microsoft 365 Copilot – 工作上的最佳 AI 助手」，The Official Microsoft Blog，2023 年 3 月 16 日，https://blogs.microsoft.com/ blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/

[8] Ruby, D.，「30 + 詳細的 ChatGPT 統計 – 使用者和事實」，《DemandSage》，2023 年 7 月 7 日，https://www.demandsage.com/chatgpt-statistics/

採用人工智慧還是不採用人工智慧

馬是長久不變的存在，但汽車只是新奇事物。一時的流行。[9]

密西根州儲蓄銀行向亨利．福特（Henry Ford）的律師霍勒斯·拉克姆（Horace Rackham）提供的建議。

有一些組織對於未知感到恐懼，選擇忽視人工智慧（AI）不斷增長的影響力。其他則害怕是另一個技術噱頭，把人工智慧（AI）的前景當作只是炒作而不再考慮。但如果運用得當，人工智慧工具可以增強人類的創意，補足業務流程，並將複雜的資訊整理成易於理解的內容，這些都能帶來更高的利潤和績效。

從車輪到汽車，歷史顯示任何讓我們事半功倍的技術都廣受歡迎，生成式AI已證實這一點。從自動化日常任務到產生創意內容，生成式AI已經展示了其提高生產力與釋放人類創造力的潛力。一些工作者已經將使用人工智慧（AI）視為正常現象，就如同查看電子郵件或線上購物一樣。人工智慧（AI）與我們的個人和職業生活深度融合，也只是時間的問題。

儘管它已滲透到社會和文化的各面向，但一些企業還是禁止或限制了人工智慧（AI）的使用。Stack Overflow 是提供程式撰寫的問答平台，它暫時禁止了 ChatGPT[10]，因為它傾向於給出聽起來自信且合理的錯誤答案，導致使用者在尋找問題的相關答案時感到困惑或被誤導。三星電子[11]、蘋果[12]、摩根大通[13]和威瑞森電信[14]出於對安全的隱憂，已嚴格限制在工作場所使用生成式AI。

這些限制是有道理的，考慮到目前人工智慧（AI）能力和後果尚不明確，但大多數企業領袖意識到，關於長期人工智慧前景的決策將遠比簡單的「是」或「否」要複雜。如果一家企業完全禁止人工智慧（AI），其競爭優勢可能受到的影響微乎其微，因為一些員工、合作夥伴、批發商和競爭對手可能會從中受益。即便企業不使用具備人工智慧（AI）的工具，忽視這項技術也可能會使企業更容易遭受風險（例如，技術過時、難以獲取頂尖人才的機會等）。

從車輪到汽車，歷史顯示任何讓我們事半功倍的技術都廣受歡迎—生成式AI已證實這一點。

因此，高階領導者如果希望利用人工智慧技術，就必須確保他們的組織擁有正確的基礎設施和治理流程。為了準確理解人工智慧（AI）所帶來的弱點和優勢，領導者必須進行全面的風險影響分析，充分考慮人工智慧領域當前的不確定性及其未來的發展潛力。

識別人工智慧的風險和報酬

人工智慧即將滲透到各行各業。從提供學生現成的學術論文到為好萊塢電影公司撰寫劇本，再到分析工廠生產線設計的瑕疵，人工智慧（AI）幾乎為所有團隊、領域和組織帶來益處。

[9] Bushnell, S.T.;「亨利·福特的真相」，The Reilly & Lee Company，美國，1992 年
[10] meta.stackoverflow.com，「暫時性政策：ChatGPT 被禁止」，https://meta.stackoverflow.com/questions/421831/temporary-poli-cy-chatgpt-is-forbidden
[11] Cawley, C;「三星在代碼洩露後限制生成式 AI 的使用」，《Tech.co》，2023 年 5 月 8 日，https://tech.co/news/samsung-restricts-genera- tive-ai-use#:~:text=New%20Policy%20Bans%20Samsung%20Employeesz
[12] Tilley, A.;M. Kruppa;「蘋果限制員工使用 ChatGPT，加入其他警惕洩密的公司」，《華爾街日報》，2023 年 5 月 18 日，https://www.wsj.com/articles/apple-restricts-use-of-chatgpt-joining-other-companies-wary-of-leaks-d44d7d34?mod=article_inline
[13] Lukpat, A.;「摩根大通限制員工使用 ChatGPT」，《華爾街日報》，2023 年 2 月 22 日，https://www.wsj.com/articles/jpmorgan-restricts-employees-from-using-chatgpt-2da5dc34?mod=article_inline
[14] Moneylife.in，「摩根大通限制工作者使用 ChatGPT」，2023 年 2 月 23 日，https://www.moneylife.in/article/jpmorgan-chase-restricts-workers-from-using-chatgpt/69948.html

在整體行業層面（以及個別企業層面），領導者必須採取四個重要步驟，在安裝適當有效的防護措施的同時，最大限度地發揮人工智慧（AI）的價值：

一、識別人工智慧的效益。

二、識別人工智慧風險。

三、採用持續的風險管理方法。

四、實施適當的人工智慧安全協議。

如果企業領導者能夠遵循這些步驟，那麼在企業中利用人工智慧工具和流程時，他們就更有可能取得風險與報酬間的良好平衡。

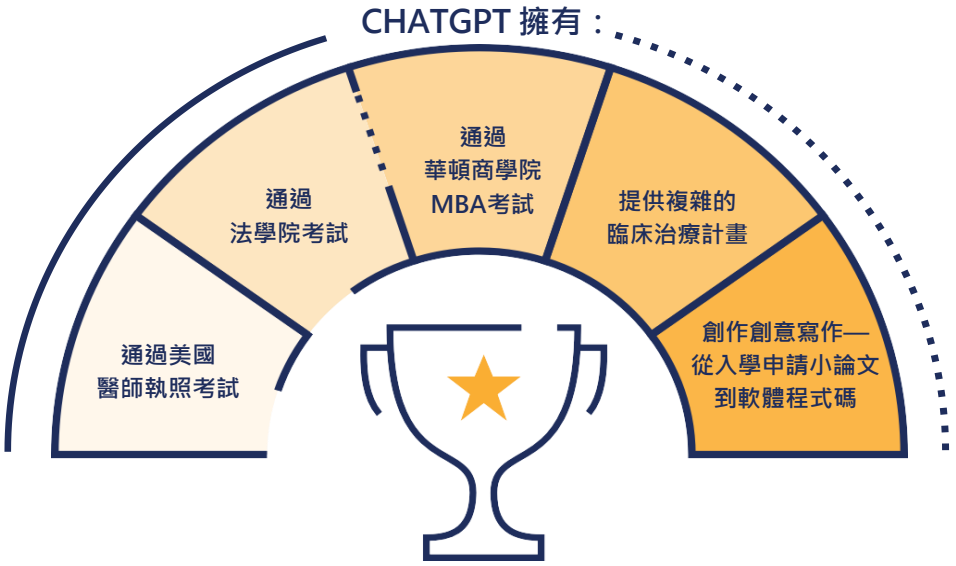
進行人工智慧效益分析

比爾·蓋茨（Bill Gates）等領導者和慈善家預見到人工智慧（AI）將帶來諸多好處，[15]從為弱勢團體提供醫療保健和教育，到矯正氣候變遷的影響。對於企業而言，潛在優勢包括更好的創新能力、提高效率、提升生產力和優化勞動力。許多公司已經在尋求使用人工智慧（AI）來實現任務自動化，並將員工重新分配到更有價值的工作中，例如，他們可以專注於設計卓越的新服務系列或發展更周到的客戶關係。

那些不願利用人工智慧（AI）的公司，很可能會發現自己走上了被淘汰的道路。正如人工智慧程式在短短幾周內吸引了數十億用戶一樣，這些程式改變市場和產品性能的速度將令人震驚。

企業巨擘可能會發現，他們使用人工智慧（AI）的競爭對手，比以往更聰明、更快速且更具威脅性。圖 2 包含一些實際使用案例。[16]

圖 2： AI 不安的糾結 — AI 比我們更聰明嗎？



[15] Gates, B.;「人工智慧時代已經開始」，《GatesNotes》，2023 年 3 月 21 日，https://www.gatesnotes.com/The-Age-of-AI-Has-Begun

[16] Ault, A.;「人工智慧機器人 ChatGPT 通過美國醫事人員執照考試不需死記硬背——與學生不同」，《Medscape》，2023 年 1 月 26 日，https://www.medscape.com/viewarticle/987549 ; Sloan, K.;「ChatGPT 儘管表現『平庸』，但還是通過了法學院考試」，《路透社》，2023 年 1 月 25 日，https://www.reuters.com/legal/transactional/chatgpt-passes-law-school-exams-despite-mediocre-performance-2023-01-25/ ;businesstoday.in，「人工智慧的下一步：ChatGPT 通過美國醫學執照考試華頓商學院 MBA 考試」，《businesstoday》，2023 年 1 月 25 日，https://www.businesstoday.in/technology/story/next-step-for-ai-chatgpt-clears-us-medical-licensing-exam-and-whartons-mba-exam-367569- 2023-01-25;tribalhealth.com，「人工智慧正在通過醫學院考試」，2023 年 1 月 24 日，https://tribalhealth.com/chatgpt/; Whitford, E.;「Forbes 如何讓 ChatGPT 人工智慧在 20 分鐘內寫出 2 篇大學論文」，《富比士》，2022 年 12 月 9 日，https://www.forbes.com/sites/emmawhitford/2022/12/09/heres-how-forbes-got-the-chatgpt-ai-to-write-2-college-essays-in-20-minutes/?sh=313d2f3b56ad ; Hutson, M;「人工智慧以『驚人的』進步學習撰寫電腦代碼」，《科學》，2022 年 12 月 8 日，https://www.science.org/content/article/ai-learns-write-computer-code-stunning-advance

人工智慧（AI）的使用加劇了競爭

A公司是行業領導者，憑藉一款旗艦產品在市場上佔據主導地位10年。執行長對公司的地位感到安全，正如競爭情報向她保證的那樣，沒有其他公司的研究和開發接近A公司產品的功能。

然後，一家名為B公司的新興競爭者，使用人工智慧（AI）來開展研究、設計突破性產品、建立定價模型和行銷計畫、推出新網站並拓展線上銷售管道。所有這些都在兩周內完成，並只用了最少的人力。

僅僅兩次員工會議之後，A公司的產品就被視為過時，其客戶群轉向B公司，利潤開始暴跌。其他新興競爭者開始使用人工智慧（AI）來推翻B公司。市場在幾個月內發生翻天覆地的變化，A公司奮力拼搏，卻最終未能保持競爭力。

要對任何技術進行全面的效益分析，領導者應該超越表面上的簡單優勢，回顧公司根本的核心價值與理念。需要考慮的因素包括：

- 公司是做什麼的，為什麼要這樣做？
- 他們是如何執行的？
- 公司的競爭對手如何善用人工智慧工具？
- 他們依賴哪些工具和人才，衡量成功的標準是什麼？
- 對人工智慧（AI）的投入將花費多少成本，它能帶來多少投資回報？
- 公司對其安全協議和機制將能夠確保資料品質、完整性和保密性的信心有多高?例如，智慧財產權(IP)、個人身份資訊(PII)等
- 當人工智慧（AI）消除了巨大的成本和人才障礙時，公司當前的策略和目標將如何演變？
- 員工和合作夥伴可以為人工智慧創新帶來什麼？

這些問題的答案將提供一個路線圖，把人工智慧（AI）使用案例與公司的使命和目標連接起來。

識別人工智慧風險

人工智慧（AI）相關的潛在風險主要被視為未來擔憂。然而，積極應對這風險有助於確保人工智慧技術在商業中安全、負責任的發展。

傑佛瑞·辛頓（Geoffrey Hinton）博士通常被稱為人工智慧（AI）的「教父」，稍早他從Google辭職，正式承認他對自己的工作感到後悔，將某些危險稱為「相當可怕」[17]。

他從事神經網路(neural networks)和深度學習(deep learning)方面工作，為我們現在稱的人工智慧程式設計奠定了基礎；他概述了未來人工智慧能力可能超出我們的理解，因此限制我們動員或設立有效保障措施的能力。

儘管意識到未來風險很重要，但人工智慧技術的現狀和我們的使用模式構成了必須解決的立即風險。

[17] Vallance, B.; C. Vallance;「人工智慧『教父』傑佛瑞·辛頓（Geoffrey Hinton）在退出谷歌時警告危險」，《BBC 新聞》，https://www.bbc.com/news/world-us-canada-65452940

這種風險可以概括為：人工智慧（AI）可能會加劇現有問題，如缺乏品質控制或資料完整性差，系統將容易受到網路攻擊，並可能引發新的攻擊。從資料洩露到CEO詐欺，人工智慧（AI）可以迅速將「受控制」風險放大到混亂的程度，並可能使毫無準備的企業偏離正軌。圖3中列出了人工智慧（AI）如何改變業務功能的一些值得注意的範例。[18]

無論人工智慧（AI）多麼有用，它並非沒有固有的風險。以下各節將介紹一些常見的風險。

社會風險

隨著人工智慧（AI）繼續影響社會，我們必須對其目標和潛在危害保持批判態度。人工智慧（AI）可以用來製造虛假資訊，以一種令人信服和破壞性的方式操縱人口和政治人口統計。人工智慧（AI）已經被用於執行新型網路犯罪（如深度偽造、複雜的網路釣魚電子郵件、人工智慧語音產生器等）。例如：五角大廈爆炸的假圖片在推特上廣泛流傳，引起股市反應。[19]

圖 3：顛覆、取代和發現——人工智慧將在多大程度上改變商業？

[18] Jones, J.;「人工智慧可以自動化 3 億個工作崗位。這是風險最大（和最小）的」，《ZDNET》，2023 年 5 月 27 日，https://www.zdnet.com/article/ ai-could-automate-25-of-all-jobs-heres-which-are-most-and-least-at-risk/; Elder, B.;高盛（Goldman Sachs）說：把你的辦公桌工作交給人工智慧生產力奇跡」，《金融時報》，2023 年 3 月 27 日，https://www.ft.com/content/50b15701-855a-4788-9a4b-5a0a9ee10561 ;Neil-Baily, M.;Brynjolfsson; A. Korinek;「思維機器：人工智慧驅動的生產力繁榮的案例」，《布魯金斯學會》，2023 年 5 月 10 日，https://www.brookings.edu/articles/machines-of-mind-the-case-for-an-ai-powered-productivity-boom/ ;Kalliamvakou, E.;「研究：量化 GitHub Copilot 對開發人員生產力和幸福感的影響」，《GitHub Blog》，2022 年 9 月 7 日，https://github.blog/2022-09-07-research-quantifying-github-copi- lots-impact-on-developer-productivity-and-happiness/ ;Korinek, A.;「經濟研究的語言模型和認知自動化」，《美國國家經濟研究局》，https://doi.org/10.3386/w30957; Brynjolfsson, E.; D. Li; L. Raymond「生成式人工智慧在工作」，《美國國家經濟研究局》，https://doi.org/10.3386/w31161; Chui, M.; R. Roberts; L. Yee;「生成式人工智慧來了：ChatGPT 等工具如何改變您的業務」，《麥肯錫》，2022 年 12 月 20 日，https://www.mckinsey.com/capabilities/quantumblack/our-insights/generative-ai-is-here-how-tools-like-chatgpt-could-change-your-business

[19] Bond,S.;「五角大廈爆炸的虛假病毒圖像可能是由 AI 建立的」，《NPR》，2023 年 5 月 22 日，https://www.npr。org/2023/05/22/1177590231/fake-viral-images-of-an-explosion-at-the-pentagon-were-probably-created-by-ai

這起事件說明了人工智慧（AI）生成的錯誤資訊可能帶來深遠的影響，破壞金融市場的穩定性並削弱對機構的信任。基於這個原因，將需要開發機制，讓企業和使用者能夠排除虛假資訊。人工智慧工具的使用者也有責任評估提供給他們的資訊，以減輕潛在的負面影響。此外，經濟動盪的可能性與失業有著直接相關；據預測，超過 3 億個工作機會[20]將受到 ChatGPT 模型的衝擊，失業更影響了許多行業和職業。

智慧財產權洩露和失效

許多人工智慧使用者欠缺考慮地將智慧財產權（IP）、商業秘密、競爭對手內容和其他資料輸入人工智慧模型，這帶來了一系列隱私風險。舉例而言。最近，三星員工通過使用 ChatGPT 洩露了公司專有資訊。[21]另一個例子可能涉及一家希望對其批發商品和定價保密的公司。

然而，如果他們允許員工使用人工智慧工具，當員工將其複製並貼到人工智慧程式中，則機密資訊可能會被公開。此外，上述所提及 Microsoft Copilot 將會搜索公司內部在 SharePoint、Excel、Slack、電子郵件、Teams messages 和 OneDrive 上的內部資料，從而分享私人和敏感數據。

因此，亞馬遜等一些公司警告員工不要與 ChatGPT 分享敏感資訊。[22]儘管如此，仍有許多公司沒有這樣的政策或警告，而其中應該包括對公司認為的敏感資訊的明確定義。

這些政策還應確保與員工保持一致，否則他們可能對自己正在與一個龐大的數位儲存庫共享的版權、商標和敏感資訊不以為意，而這些資料庫可能會將這些資訊分享給其他人。

還應該提到的是，心懷不滿的員工意識到分享這些資訊的潛在危害，可能會故意向人工智慧工具提供資料，從而損害公司的聲譽。由於智慧財產權保護受到複雜法律的約束，員工在向人工智慧應用程式中輸入資訊，可能會使專利和其他智慧財產權失效。例如，如果一家公司因為員工的疏忽而未能充分保護其營業秘密，則可能會喪失對保密資訊的法律保護。

擁有權無效

任何使用人工智慧（AI）創造產出的，如行銷用的品牌理念，公司都必須保證它是真正的原創智慧財產權，而不是他人作品的衍生品。[23]否則，這些企業可能會發現他們對新資產的所有權是無效的。此外，美國著作權局已經拒絕了一些人工智慧（AI）生成的圖像的版權，[24]這取決於它們是根據文字提示詞建立的，還是反映了創作者「自己的思維概念」。

例如，如果作者複製並貼上一個令人欽佩競爭對手內容的連結作為人工智慧工具的靈感，並回饋半抄襲的內容[25]，則也可能發生所有權無效的情況。關於人工智慧（AI）的無效所有權主張最終可能導致法律糾紛，並對企業的聲譽產生不利影響。

[20] 同前文所引 Jones
[21] Maddison,L.;「三星員工使用 ChatGPT 犯了一個重大錯誤」，《Techradar》，2023 年 4 月 4 日，https://www.techradar.com/news/samsung-workers-leaked-company-secrets-by-using-chatgpt
[22] Sengupta,A.;「亞馬遜警告員工注意 ChatGPT，不要與聊天機器人分享敏感資訊」，《今日印度》，2023 年 1 月 27 日，https://www.indiatoday.in/technology/news/story/amazon-warns-employees-chatgpt-do-not-share-sensitive-info-with-chatbot-2327014-2023-01-27
[23] McKendrick ,J.;「誰最終擁有 ChatGPT 和其他 AI 平台生成的內容？」，《富比士》，2023 年 2 月 22 日，https://www.forbes.com/sites/joemckendrick/2022/12/21/who-ultimately-owns-content-generated-by-chatgpt-and-other-ai-platforms/?sh=21cf97855423
[24] Brittain,B.;「美國版權局表示，一些人工智慧輔助的作品可能受版權保護」，《路透社》，2023 年 3 月 15 日，https://www.reuters.com/world/us/us-copyright-office-says-some-ai-assisted-works-may-be-copyrighted-2023-03-15/
[25] Duffy, J.;「為什麼作家知道使用 ChatGPT 是個壞主意」，《PCMAG》，2023 年 1 月 25 日，https://www.pcmag.com/opinions/why-writers-know-using-chatgpt-is-a-bad-idea

資通安全和韌性衝擊

從應用程式介面（API）整合到創建越來越有說服力的網路釣魚電子郵件，人工智慧（AI）為一個更加複雜的網路犯罪世界打開了大門。行為惡劣的人已經在使用人工智慧（AI）來更快地編寫惡意軟體，生成駭客腳本，發起勒索軟體攻擊，並逼真地模仿CEO的聲音。[26]

人工智慧（AI）的易用性可能是其最大風險，因為攻擊者需要做的就是在人工智慧工具中插入一個簡單的指令。幾乎不需要程式設計或專業知識即可發動攻擊。這使得人工智慧（AI）的使用大眾化，幾乎所有個人都能利用其力量進行破壞。

人工智慧（AI）為一個更加複雜的網路犯罪世界打開了大門。

人工智慧（AI）為缺乏技術技能的潛在攻擊者，提供了進入網路犯罪的階梯，因為他們可以簡單地使用人工智慧（AI）驅動的滲透預測試工具，如PentestGPT，以及「為我找到特定技術的安全漏洞，像是Windows、亞馬遜網路服務或工業控制系統，並編寫代碼來利用它」之類的提示詞。然後，攻擊者可以從電子郵件伺服器貼上程式碼，人工智慧（AI）將編寫腳本來利用該漏洞。

備份和災難復原是另一個不應該被忽視的問題。從人工智慧相關事件以及人工智慧驅動的回應計畫的觀點來看，都必須將人工智慧（AI）納入事件回應和營運持續計畫中。

PentestGPT

- ChatGPT 已被用於建立資訊竊取程式、加密工具和暗網惡意軟體腳本。[27]
- Darktrace報告稱，在 1 月至 2 月期間，發送給客戶的垃圾郵件增加了135%[28]，其特點是英語文法和語法明顯更好。該公司表示，他們認為駭客使用生成式AI應用程式來策畫他們的攻擊活動，聽起來更有說服力的美式風格。[29]
- Zscaler 執行長傑·喬杜里的聲音影片片段，被人工智慧工具轉化為非常有效的 CEO 欺詐嘗試[30]。該公司還表示，人工智慧（AI）是 Zscaler 於 2022 年執行網路釣魚攻擊增加 47% 的一個因素。[31]
- 根據黑莓研究公司的資料，51%的資訊專業人士預測，到今年年底我們將見證在ChatGPT[32]的幫助下成功發動的網路攻擊。

內部權限結構薄弱

許多商業領袖在人工智慧（AI）中發現的第一個機會是優化內部系統，如企業資源規劃或定價模型和庫存系統。他們沒有預見到，在為這些系統輸入資訊後，員工能夠向同一工具詢問同事的薪資和其他敏感資訊。雖然這種風險似乎並非人工智慧（AI）所獨有，但考慮到使用PassGPT（一種用於猜測和生成密碼的LLM）[33]等工具如何顯著增加存取漏洞和其他安全漏洞的可能性。

[26] Menn, J.;「網络安全面臨人工智慧崛起的挑戰」，《華盛頓郵報》，2023 年 5 月 11 日，https://www.washingtonpost.com/technology/2023/05/11/hacking-ai-cybersecurity-future/
[27] research.checkpoint.com，「OPWNAI：網路犯罪分子開始使用 ChatGPT」，2023 年 1 月 6 日，https://research.checkpoint.com/2023/opwnai-cybercriminals-starting-to-use-chatgpt/
[28] darktrace.com，「Darktrace/Email™ 產品的重大升級可保護組織免受不斷發展的網路威脅環境，包括生成式人工智慧商業電子郵件洩露和新型社會工程攻擊」，2023 年 4 月 3 日，https://darktrace.com/news/darktrace-email-defends-organizations-against-evolving-cyber-threat-landscape
[29] 同上。
[30] foxbusiness.com，「Zscaler 的 Jay Chaudhry 給出了人工智慧深度偽造的令人不寒而慄的例子」，2023 年 6 月 2 日，https://www.foxbusiness.com/video/6328691103112
[31] Zscaler.com，「Zscaler ThreatLabz 2023 網络釣魚報告」，https://info.zscaler.com/resources-industry-reports-threatlabz-phishing-report
[32] Singh,S.;「IT 領導者預測 ChatGPT 支援的網络攻擊迫在眉睫」，《Blackberry》，2023 年 2 月 2 日，https://blogs.blackberry.com/en/2023/02/it-leaders-predict-chatgpt-enabled-cyberattacks-are-imminent
[33] Lanz, J. A;「認識 PassGPT，一個在數百萬個洩露密碼上訓練的人工智慧」，《Decrypt》，2023 年 6 月 9 日，https://decrypt.co/144004/meet-passgpt-ai-trained-millions-leaked-passwords

技能差距

2023年3月，拜登·賀錦麗(Biden-Harris)政府宣佈了新的《國家資通安全戰略》。[34]其中一些影響深遠的變化，包括重新平衡安全控制的責任，並要求軟體供應商承擔更多的責任。為了切實實施這些措施，資訊人員將需要先進的技能，但他們目前可能不具備此技能來保護企業免受高技能的人工智慧攻擊。這也意味著對現有員工進行再培訓或招募必要人才的預算可能會產生重大影響。

網路專家還敦促建立一個全國性的數位身份框架[35]，以追蹤公民活動。實施和控制這些變革需要超越一般資訊部門的先進技術技能。

反應過度

資訊專業人士長期以來一直觀察到在勒索軟體攻擊後的模式。忽視資通安全計畫的企業會受到攻擊，之後他們會驚慌失措並花費過多資金於安全工具上，但這些工具並不總是相輔相成的，這可能會增加他們遭受第二次攻擊的風險。

隨著企業意識到人工智慧安全漏洞的範圍，或者自己經歷災難，並在應對中採取過度的措施時，同樣的場景可能會上演。

雖然某些企業會將人工智慧（AI）排除在業務流程之外，但其他企業可能會反其道而行，對人工智慧（AI）反應過度，並過度信任未經驗證的人工智慧輸出。員工可能會發現他們承擔了難以承受的工作量，於是便採取變通方法，而沒有對人工智慧的輸出進行測試。這些未經測試的假設將加劇品質控制問題，並給組織帶來聲譽風險。

預期和非預期用途

目前大多數人工智慧應用程式在其設計中都嵌入了道德限制，以防止濫用其知識。例如，由於ChatGPT 等工具內建了護欄，詢問人工智慧（AI）如何實施暴力行為的人，可能會被建議去看心理醫師。但不幸的是，使用者也已經找到了一種方法來利用越獄[36]提示或使用WormGPT[37]等工具來繞過這些限制，導致商業電子郵件入侵（BEC）攻擊的複雜程度增加，攻擊者可以力入這些攻擊獲取從製造武器到實施大規模暴力等各種主題的資訊。

資料完整性

ChatGPT、Bard 和其他程式往往會產生有嚴重缺陷的內容，這些內容看起來正當且聽起來很自信。這導致使用者對生成的輸出過於信任。當人工智慧（AI）產生錯誤資訊或沒有現實世界數據支援的資訊時，它被稱為「幻覺」。[38]使用者還注意到，不同的人工智慧程式會為歷史問題提供不同的答案[39]，並提供沒有引用來源的資訊。

ChatGPT的創辦人兼Open.AI執行長Sam Altman 在推特[40]上承認，「現在依賴『AI 工具』做任何重要的事情都是錯誤的」，並補充說:「我們在穩健性和真實性方面還有很多工作要做。」

[34] Whitehouse.gov，「概況介紹：拜登-賀錦麗政府宣佈國家網路安全戰略」，2023 年 3 月 2 日， https://www.whitehouse.gov/briefing-room/statements-releases/2023/03/02/fact-sheet-biden-harris-administration-announces-national-cybersecurity-strategy/

[35] Macdonald, A.;「拜登敦促考慮聯邦數位身份框架」,《Biometric Update》, 2023 年 8 月 1 日, https://www.biometricupdate.com/ 202204/biden-urged-to-consider-federal-digital-identity-framework

[36] Loynds, J.;「如何越獄 ChatGPT：最佳提示及更多」,《Dexerto》, 2023 年 8 月 1 日, https://www.dexerto.com/tech/how-to-jailbreak-chatgpt-2143442/

[37] Kelley, D.;「WormGPT - 網络犯罪分子用來發動 BEC 攻擊的生成式 AI 工具,」,《SlashNext》, 2023 年 7 月 13 日, https://slashnext.com/blog/wormgpt-the-generative-ai-tool-cybercriminals-are-using-to-launch-business-email-compromise-attacks/

[38] en.wikipedia.org,「幻覺（人工智慧）」, https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence)#%3A~%3Atext%3DFor%20example%2C%20a%20hallucinating%20chatbot

[39] McCracken, H.;「如果 ChatGPT 不能更好地掌握事實，其他一切都無關緊要」,《Fast Company》, 2023 年 1 月 11 日, https://www.fastcompany.com/90833017/openai-chatgpt-accuracy-gpt-4

[40] Altman, S.;「ChatGPT 的局限性令人難以置信，但在某些事情上足夠出色，足以給人留下偉大的誤導性印象。 現在依賴它來做任何重要的事情都是錯誤的。這是進展的預覽;在穩健性和真實性方面，我們還有很多工作要做。」,《 Twitter》, 2022 年 12 月 10 日, https://twitter.com/sama/status/1601731295792414720?lang=en

因此，企業應考慮確保人類是最終的決策者，對用於做出影響生命決策的產出實施雙重控制，例如醫療、刑事司法、監控、監禁等。在這種情況下，具備人工智慧（AI）的工具可以提出建議，但不負責最終決定。

有些使用者不知道如何使用正確的 ChatGPT 提示，導致來自不良來源的資料輸出不完整或不正確的資料。此外，由於缺乏透明度和可追溯性，使用者可能不信任用於訓練人工智慧模型的語料庫（即資料）。為此，公司如何檢測剽竊，評估資料的權威性和來源，或者識別資料或人工智慧演算法中內建的系統性偏見？當處理的資訊量過大到任何人類都無法處理時，這將很難證明（並且幾乎不可能檢測到）。

資料的可靠性取決於來源。目前，人工智慧工具無法保證資料的完整性。這種不確定性讓使用者難以判斷哪些資訊是可信的，哪些資訊應該被驗證或忽略。例如：一名律師處理一份預計需要50到55小時的法律簡報，可能發現生成式AI工具可以在不到5分鐘的時間內完成簡報。但是，由於資訊量實在是太大，沒有時間審查和評估工作的品質，律師可能不再進行驗證其真實性。這種情況於 2023 年 5 月在曼哈頓聯邦法院出現。[41]

如果沒有適當的資料完整性措施，員工使用人工智慧工具可能會產生不正確或誤導性的結果，這可能會對企業及其資料品質產生嚴重後果。

責任

當一個人犯罪或在工作中犯下重大錯誤時，究責和賠償是毫無疑問的。但是，例如，人工智慧聊天機器人對未成年人發表不當言論，誰應該承擔責任？訴訟將針對誰？參與設計人工智慧（AI）的工程師是犯罪的共犯嗎？誰應該負責制定有關人工智慧（AI）使用和問責的決策？提出這些問題並討論其影響非常重要，因為利用人工智慧（AI）來尋求可行商業市場的公司，最終可能會承擔意想不到的後果，而沒有任何補救的理由。

這將成為工作自動化的首要考慮因素。任何人類員工都被賦予了一定程度的信任和責任，但是當那個人被機器取代時，如果機器出現問題，誰應該承擔責任？當人工智慧（AI）做錯事時，由員工引起的個人責任問題就會變成公司責任問題。在過去，可能通過糾正員工的行為來解決違規行為。然而，當沒有人可以被追究責任時，誰來承擔責任？

任何人類員工都被賦予了一定程度的信任和責任，但是當那個人被機器取代時，當它出錯時，誰應該承擔責任？

當人工智慧風險資訊被媒體報導或廣為人知時，新的風險可能已被發現。我們鼓勵從業人員藉由追蹤學術界和法規的最新動態，密切關注人工智慧（AI）的最新發展。

[41] Weiser, B.;「當你的律師使用 Chat GPT 時會發生什麼」，《紐約時報》，2023 年 5 月 27 日，https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html

採用持續風險管理方法

採用框架可以幫助剛開始開發持續風險評估方法的團隊。例如：美國國家標準暨技術研究院（NIST）的人工智慧風險管理框架（AI RMF 1.0）[42]等新框架，旨在引導企業進入未知的人工智慧（AI）領域。

重要的是要記住，接受一些風險是健康且有利的，而完全規避風險，尤其是在從新興技術中受益時，可能會付出過高的成本。為了取得適當的平衡，採用持續風險方法的實施包括三個步驟：

一、識別公司的整體人工智慧風險。

二、定義公司的風險胃納。

三、監控和管理風險。

第一步：識別風險

為了制定量身定製的風險格局，高階管理階層和員工應該研究可能的損失事件情景及其對組織目標的衝擊。使用調查、訪談、工作小組和腦力激盪會議來闡明全方位的見解。圖4 顯示了風險識別流程範例。

第二步：定義風險胃納

應根據每種風險的潛在影響嚴重程度和發生的可能性進行評估並排定其優先順序。為了幫助促進這一點，可以成立一個人工智慧探索小組委員會，負責向企業風險管理階層報告調查結果和建議。

圖4：風險識別方法

制訂使命和目標

- 作為一個組織，您如何為利益相關者和客戶提供價值？
- 價值交付是按照部門別還是業務管理層別來分組？

確定服務和產品

- 哪些服務和產品可以實現使命和目標？

確定路徑和資產

- 什麼將幫助於您提供產品和服務？
- 哪些技術、業務流程、員工、廠商、供應商和資料對於生產您的服務和產品至關重要？

建立人工智慧價值鏈

- 在價值鏈的哪個方面可以部署人工智慧（AI）來維持或增強競爭優勢？
- 是什麼給公司/組織帶來了競爭優勢，以及如何？

評估 AI 影響

- 人工智慧（AI）如何影響使命、目標、服務、產品和關鍵資產？
- 人工智慧（AI）給組織帶來了哪些威脅和好處？

[42] NIST，「人工智慧風險管理框架（AI RMF 1.0）」，2023 年 1 月，https://doi.org/10.6028/nist.ai.100-1

該委員會由關鍵利害關係人組成，負責推動必要的工作，以確定哪些風險符合企業的風險參數，哪些風險需要額外的控制才能融入企業的文化和目標。有些風險將被評估為具有過高的潛在成本。

一旦確定了風險胃納，就應該將它記錄並分享，以確保每位團隊成員理解他們在遵守這些限制中的作用。

第三步：監控和管理風險

為了監控和管理風險，領導層必須認識到人工智慧風險對其企業的潛在衝擊，根據業務目標確定此類風險的優先順序，並採取必要措施減輕風險，同時實現價值最大化。對於第三步，由業務和技術利害關係者組成的跨域監督團隊（包括法律、風險與法遵），應該啟動風險管理流程。

此流程包括以下內容：

- 向所有員工傳達風險願景
- 確定風險管理活動的優先次序
- 解決需要採取行動的所有風險（例如：新的安全控制），並指導團隊成員解決其風險領域
- 分配所有權和個人責任
- 確定風險監控的頻率

一旦風險管理過程順利運行，風險監督團隊就應該衡量其新安全控制的成功與否，追蹤風險計畫的偏差，並繼續評估新風險。

正如人工智慧（AI）一直在發展一樣，風險管理也是持續進行。因此，企業必須採用持續的風險方法框架，頻繁和精確測試假設，以確保其人工智慧（AI）輸出的品質。

建構人工智慧安全計畫：八種協議和實踐

有效的網路安全始終需要採積極主動的方法，對於人工智慧（AI）來說尤其如此。建立風險管理基礎之後，企業必須啟動安全策略和控制措施，以降低資料洩漏和其他人工智慧（AI）濫用的可能性。

小型企業可能會覺得其中一些控制和做法超出了他們的範圍。但事實上，對於可能沒有資安長(CISO）或專門網路團隊的小型企業來說，人工智慧（AI）可以成為一個巨大的安全優勢，因為它可以透過自動化流程審查或這些人力資源日常負責的非關鍵決策。或者，通過聘請具有專業人工智慧知識的合作夥伴，這些企業可以在不增加員工人數或超出預算的情況下加強其安全態勢。

本節中討論的八種做法可能無法解釋人工智慧（AI）未來發展和脆弱性，但它們可以為任何開始人工智慧之旅的企業提供一個全面的保護基礎。

一：信任但要驗證

在人工智慧革命的早期，執法機關已經使用人工智慧工具來預測犯罪，金融機構也使用人工智慧程式來預測貸款違約。顯然，人工智慧（AI）的潛力具有提高效率、增加資料導向實踐和擴展特定任務和決策能力，從而使執法、金融和許多其他業務部門能夠做出更明智、更有效的決策。組織利用人工智慧（AI）可能會進行驗證和盲目地

遵循第一次人工智慧（AI）生成的輸出。這就是組織可能會對人工智慧準確性產生錯誤信任的地方，就像我們相信數學計算機是正確的並且不再費心去檢查一樣。

然而，重要的是要記住，人工智慧（AI）不是一個計算機；它是一個不斷發展的複雜工具，從一大堆模糊的資料源中提取。由於其反覆運算性，人工智慧平臺將會不斷進行升級，同時也會遭到駭客攻擊和滲透。當 OpenAI 在2023年5月確認其系統中存在資料洩露時，我們已經看到了這一點。[43]安全研究人員、技術人員和業餘愛好者已經開始建立「越獄」，以繞過人工智慧（AI）[44]的規則，產生不想要的內容，甚至將惡意軟體放入人工智慧模型本身。因此，必須持續驗證所有輸出。這項可能繁重的任務，要求團隊開發機制來評估和批准所有人工智慧（AI）生成的工作。

二：制定可接受的使用政策

目前，大多數員工可以使用人工智慧（AI）來設計橫幅廣告、撰寫年度報告或檢查銷售模式的競爭力。人工智慧（AI）的這種使用帶來了幾個問題，包括共用公司專有資料和存取與其工作職責不相稱的資訊。政策領導者必須制定程式和規則，以強制執行安全和合乎道德的使用人工智慧（AI）。他們還必須與不同的利害關係人和主題專家合作，設計這些政策並對其進行壓力測試，以防止無意的偏見和歧視。政策應加強或修改，以確保它們符合其營運所在司法管轄下不斷變化的外部要求（例如，《歐盟人工智慧法案》）。

此外，員工應接受有關適當和不適當使用這些工具的訓練，以降低風險。降低風險的另一種方法是實施核准鏈和審查流程，這有助於預防內部威脅。

三：指定人工智慧主管

未來，許多企業的高層管理階層可能會設置一個人工智慧主管。現在這可能還為時過早，但組織應該考慮指派一名分析師或專案經理，以追蹤人工智慧（AI）的發展並制定專屬計畫，記錄公司與人工智慧工具不斷發展的關係。重要的是，人工智慧主管要與跨職能和多元化的利害關係人協作，其中至少包括來自資通安全、資料隱私、保密、法律、採購、風險和稽核部門的代表。

正如我們所見，企業對人工智慧（AI）的應用在短短幾個月內就以閃電般的速度發展。為了更好地引導智慧化AI的使用，企業應該開始記錄使用人工智慧（AI）的歷史，以追蹤錯誤、資源浪費和被忽略的機會。持續追蹤這段歷史還將有助於確保人工智慧導向的決策可以得到解釋，因為實施步驟將是可重複的，從而提高透明度。

四：執行成本分析

與實施資通安全措施類似，利用人工智慧（AI）可能會給組織帶來財務負擔。對人工智慧（AI）進行徹底的成本效益分析，例如：評估是否自建或購買人工智慧工具，將是關鍵的第一步驟，這些決定將隨著成本限制的降低而持續演進。最近，史丹佛大學的一個學生團隊僅花費了$600美元[45]就構建了一個名為 Alpaca 的人工智慧平台，證明了具有成本效益的人工智慧解決方案是行的。組織不僅需要計算人工智慧（AI）作為工具的成本效益，還需要計算安全控制的價格以及可能的生產力提升和勞動力優化。

[43] Poremba，S.;「ChatGPT 確認資料洩露，引發安全問題」，《安全情報》 https://securityintelligence.com/articles/chatgpt-confirms-data-breach/
[44] Burgess, M.;「ChatGPT 的駭客攻擊才剛剛開始」，《連線》，2023 年 4 月 13 日， https://www.wired.com/story/chatgpt-jailbreak-generative-ai-hacking/
[45] Wodecki, B.;「遇見 Alpaca：開源 ChatGPT 售價不到 600 美元」，《AI Business》，2023 年 3 月 20 日， https://aibusiness.com/nlp/meet-alpaca-the-open-source-chatgpt-made-for-less-than-600

五：適應和建立資通安全計畫

資安長(CISO)及其安全團隊在保護企業方面有兩項重要措施：適應當前的網路計劃，並在整個組織中建立以人工智慧（AI）為中心的新安全實踐（例如，安全設計）。這一過程應該在對人工智慧策略進行任何投資之前儘快開始。

透過利用以往的風險評估工作，企業可以對人工智慧（AI）的使用實施控制。這將有助於保護他們免受其他準備不足的組織中不可避免的安全事件的影響。

確保與人工智慧（AI）相關的風險考量和安全解決方案不是事後才想到的，可以減少昂貴的技術解決方案重做或重新設計。

以人工智慧為中心的安全實踐

有82%的資訊專業人士[46]會考慮在未來兩年內投資資通安全，以保護他們的組織免受人工智慧（AI）增強的資通攻擊，其中48%的人今年會考慮這樣做。

有95%的參與專業人士認為政府在監管此類型的技術負有一定的責任，其中85%的人認為這種責任等級為「中等」或「重大」。

在建立人工智慧資通安全計畫時，需要考慮的幾個面向包括：

智慧財產權 (IP)洩漏：安全和隱私團隊應具有適當的技能和控制措施，以防止智慧財產權洩漏。組織可以使用基於權限的存取、可視性工具以及防火牆和應用程式控制等措施，來確定員工在何處以及如何使用人工智慧工具。這些措施可以防止未經授權移轉有價值的公司資料，並限制潛在的損害。例如：當對資料建立適當以角色為基礎的存取控制時，人資經理可能能夠向人工智慧工具詢問大型語言模型(LLM)中的薪資，但只有公司律師可以詢問未結案的人資案件。此外，數位篩選器可以阻止使用者存取私有內部儲存庫。

透過教育訓練和簽署保密承諾書也可以使員工瞭解與人工智慧工具共用機敏資訊的嚴重後果。

- **災難復原和事件回應：**人工智慧（AI）可以通過多種方式協助營運持續性管理和事件回應規劃。透過分析營運數據，人工智慧工具可以識別對營運持續性的威脅，協助制定回應策略，並在模擬場景中進行測試。當發生意外造成服務中斷時，人工智慧（AI）可以自動執行營運復原計畫的某些階段，例如:優化資源分配以保持關鍵營運的正常運作。
- **持續性：**在組織內使用人工智慧（AI）會給營運持續性帶來獨特的挑戰。如果人工智慧工具停止運行，嚴重依賴人工智慧（AI）進行業務流程和功能的組織可能會發現自己無法營運。因此，在制定人工智慧計畫和策略時，必須特別考慮營運持續性。
- **威脅情報：**雖然市場上的許多威脅情報和網路監控工具，已經以某種方式整合了機器學習，但最新的生成式AI工具可以提供有關輸出的新見解，甚至可以檢測以前遺漏的威脅警報。Google發布的Google Cloud Security AI Workbench[47]，它可以分析程式碼、邊界安全和網路中的漏洞。該平臺建立個人化的威脅概況，透過可操作的情報和跨資料源的可視性，幫助團隊主動加強防禦。實施此類工具的組織可能會從威脅意識和補救措施效率的提高中受益。

六：強制進行稽核和可追溯性

企業將需要圍繞人工智慧模型提供更好的稽核和可追溯性能力，以瞭解人工智慧工具將資料從何處獲取，以及它如何做出決策。資料源來自哪裡？這些資料是否由人工智慧織免受人工操縱或是與它互動的人類所操縱？系統性偏見是否為一個因素？這些問題在評估人工智慧工具的可信度時是不可或缺的。對更值得信賴的人工智慧工具的需求，可能會成為人工智慧模型之間的競爭因素，平臺可以透明地提供模型是如何做出決策。

[46] Singh, S;「IT 領導者預測 ChatGPT 支援的網絡攻擊迫在眉睫」，《BlackBerry》，2023 年 2 月 2 日，https://blogs.blackberry.com/en/2023/02/it-leaders-predict-chatgpt-enabled-cyberattacks-are-imminent

[47] Potti,S.;「Google Cloud 計劃如何利用生成式 AI 增強安全性」，《Google Cloud Blog》，2023 年 4 月 14 日，https://cloud.google.com/blog/products/identity-security/rsa-google-cloud-security-ai-workbench-generative-ai

七：制定一套人工智慧倫理規範

許多專業人士發現，人工智慧（AI）可以幫助他們在原本所需時間的一小部分內，交付服務或產品。但是，如果他們按小時計費，則必須重新評估其定價模型和客戶合約。他們還需要注意微軟(Microsoft)要求揭露人工智慧（AI）互動（例如：ChatGPT建立的作品）。[48]

許多人可能會忽略該指令。如果平面設計師按小時向客戶收費，並報價這個專案需要15個小時的工作時間，他們需要揭露人工智慧（AI）的使用情形。但這可能暴露出他們在這個專案上只花了1個小時，而非15個小時。因此，人工智慧（AI）的使用在道德面向上帶來了許多潛在挑戰。新軟體可以檢測人工智慧（AI）的創造力，就像學術誠信軟體可以檢測作弊和抄襲一樣，但未必能發現到所有問題。每個組織和行業都必須將人工智慧倫理問題作為業務運作的嚴格對待標準。

人們已經開始關注人工智慧（AI）在我們日常生活中的倫理層面。聯合國教育、科學及文化組織（UNESCO）致力於在許多科學和技術領域建立和執行倫理護欄，並於2021年11月就出版了人工智慧（AI）的倫理問題的建議[49]。IBM[50]和美國國防部[51]等組織已經採用了人工智慧倫理原則，但我們需要共同努力，就指導方針達成並確保它們得到一致的執行。

「在人工智慧領域，沒有哪個領域的倫理羅盤比這更重要……人工智慧技術在許多領域帶來了重大利益，但如果沒有道德護欄，它就有可能重現現實世界的偏見和歧視，助長分裂並威脅到基本人權和自由。」[52]

—加芙列拉·拉莫斯（Gabriela Ramos），聯合國教科文組織社會與人文科學助理總幹事

八：社會適應

許多組織的決策將重大影響人工智慧（AI）對經濟的衝擊。隨著人工智慧工具的引入，面臨勞動力大規模失業的組織，可以選擇幫員工進行培訓，並轉型至不同工作提供新收入來源。

從中學教師到醫學院學生，學術界可能不得不改變他們的評估方法，而工作場所可能需要重新考慮他們的績效評估標準。兒童和成人都需要接受教育，瞭解深度偽造、人工智慧（AI）錯誤和虛假資訊活動的現實，正如我們目前努力教育公民有關網路詐騙的知識。

此外，在員工選才和安置過程中使用基於人工智慧（AI）的評估工具，可能會導致不合格的候選人入職。因此，需要重新評估這些過程，以降低與使用人工智慧（AI）相關的風險。

[48] Microsoft 負責人工智慧標準，v2 外部發佈的一般要求。（2022）. https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5cmFl
[49] unesco.org;「人工智慧倫理學」，https://www.unesco.org/en/artificial-intelligence/recommendation-ethics
[50] ibm.com;「什麼是人工智慧倫理」，https://www.ibm.com/topics/ai-ethics
[51] defense.gov;「國防部採用人工智慧的道德原則」，2020 年 2 月 24 日，https://www.defense.gov/News/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/
[52] UNESCO;「教科文組織關於人工智慧倫理的建議書：關鍵事實」，2023 年，https://www.unesco.org/en/articles/unescos-recommendation-ethics-artificial-intelligence-key-facts

在人工智慧驅動的未來中繁榮發展

柏拉圖在《理想國》中說：「開始是工作中最重要的部分。」組織如何開始對人工智慧（AI）進行應對之道，將決定我們是否能長期利用其優勢並抵禦其風險。作為一個社會，現在是時候回顧過去，反思並規劃人工智慧革命的風險和影響了。

市場的飛速發展，以及對監管的呼籲，或許會讓領導者在正式採用人工智慧（AI）或制定策略之前有所顧慮。但與許多技術的情況一樣，用戶已經躍躍欲試，掀起了一場需要立即採取行動的運動。分析和行動刻不容緩，行動迅速的組織將在這個人工智慧（AI）新時代創造最堅實的基礎，並取得最大的發展。

致謝

開發領袖

Ryan Cloutier
CISSP
President, Security Studio, USA

審校專家

Urs Fischer
CISA, CRISC, CIA, CPA (Swiss)
Director, Tech Cyber Security Specialist, UBS Business Solutions AG, Switzerland

Larisa Gabudeanu
CDPSE, CIPM, CIPP/E
Researcher, Babes Bolyai University, Romania

Maria Koslunova
CIPM, CIPP, FIP
Data Privacy Educator, York University, UK

David Kuo
CISA, CIPT
Distinguished Fellow, Ponemon Institute, USA

Carol Lee
CISM, CRISC, CDPSE, CCISO, CCSP, CEH, Certified Change Management Practitioner, CIPM, CSSLP
Head of Cyber Security & Risk Management, Hang Lung Properties Ltd, China

Nsuhoridem Ndeokwelu
CGEIT, CRISC, CDPSE
IT Risk and Compliance Lead, Central Bank of Nigeria, Nigeria

Dina Nu'man
CRISC, ISACA COBIT Lead Assessor, ITIL Head of Advanced Governance & Management, Scanwave, Jordan

Dilek Ozdemirci
CMMI Instructor, CMMI Lead Appraiser, CSM, PMP, SAFe PC, ITIL
Process Improvement Consultant, Dora Process Consulting Inc., Canada

ShanShan Pa
CISA, CISM, CDPSE, CIPM, CIPP/E, CIPP/ US, CIPT, FIP
Managing Director, Alpha Technology Risk Management, State Street, USA

Max Shanahan
CISA, CGEIT, FCPA, MACS (senior), MIIA
(Aust)
Governance and Assurance Consultant, Self Employed, Australia

Greg Shields
CISA, CRISC, CDPSE, CIPM, CIPT, CISSP
Senior Manager, Confidentiality and Privacy, Deloitte, USA

ISACA 董事會

John De Santis, Chair
Former Chairman and Chief Executive Officer, HyTrust, Inc., USA

Brennan P. Baybeck, Vice-Chair
CISA, CISM, CRISC, CISSP
Senior Vice President and Chief Information Security Officer for Customer Services, Oracle Corporation, USA

Stephen Gilfus
Managing Director, Oversight Ventures LLC, Chairman, Gilfus Education Group and Founder, Blackboard Inc., USA

Niel Harper
CISA, CRISC, CDPSE, CISSP, NACD.DC
Chief Information Security Officer, Data Privacy Officer, Doodle GmbH, France

Gabriela Hernandez-Cardoso
NACD.DC
Independent Board Member, Mexico

Jason Lau
CISA, CISM, CGEIT, CRISC, CDPSE, CIPM, CIPP/E, CIPT, CISSP, FIP, HCISPP
Chief Information Security Officer, Crypto. com, Singapore

Massimo Migliuolo
Independent Director, Former Chief Executive Officer and Executive Director, VADS Berhad Telekom, Malaysia

Alok Tuteja
CGEIT, CRISC
Global Head of Governance Risk and Compliance, Agthia PJSC, UAE

Patricia Voight
CISA, CISM, CGEIT, CRISC, CDSPE
Managing Director, FSO Consulting–Technology Risk, Ernst & Young LLP, USA

Maureen O'Connell
NACD.DC
Board Chair, Acacia Research (NASDAQ), Former Chief Financial Officer and Chief Administration Officer, Scholastic, Inc., USA

Erik Prusch
Chief Executive Officer, ISACA, USA

Asaf Weisberg
CISA, CISM, CGEIT, CRISC, CDPSE, CSX-P
Chief Executive Officer, introSight Ltd., Israel

Pamela Nigro
ISACA Board Chair 2022-2023
CISA, CGEIT, CRISC, CDPSE, CRMA Vice President, Security, Medecision, USA

Gregory Touhill
ISACA Board Chair 2021-2022
CISM, CISSP
Director, CERT Center, Carnegie Mellon University, USA

Tracey Dedrick
ISACA Board Chair, 2020-2021
Former Chief Risk Officer, Hudson City Bancorp, USA

ISACA 介紹

ISACA®（https://www.isaca.org/）是一個推動個人與組織追求數位信任的全球性社群。 50 多年來，ISACA 為個人與企業提供知識、證書、教育、培訓與社群，以促進其職涯發展、改造其組織並建立一個更值得信賴及合乎倫理道德的數位世界。 ISACA 是一個全球性專業協會與學習型組織，擴展其在資訊安全、治理、確信、風險、隱私及品質等數位信任領域工作的 180,000 名成員的專業技能。它已在全球 188 個國家共設立225 個分會。 ISACA 透過其 One In Tech 基金會，為資源不足及代表性不足的人們提供 IT 教育與職涯發展途徑之支援。

1700 E. Golf Road， Suite 400
Schaumburg， IL 60173， UAS

電話： +1.847.660.5505

傳真： +1.847.253.1755

技術支援： support.isaca.org

網站： www.isaca.org

參加 ISACA 線上論壇：
https://engage.isaca.org/ onlineforums

Twitter:
www.twitter.com/ISACANews

LinkedIn：
www.linkedin.com/company/isaca

Facebook：
www.facebook.com/ISACAGlobal

Instagram：
www.instagram.com/isacanews/

免責聲明

權利保留

中文版致謝名單

ISACA台灣分會 高進光理事長

翻譯: 莊盛祺

校稿: 李逸元、陳政龍

編輯排版: 游恬欣