人工智能可化身黑客 自主進行網絡攻擊

Anthropic新模型爆安全風險 銀行業警戒

美國人工智能（AI）公司Anthropic再惹爭議。該公司於7日推出最新AI模型「Claude Mythos Preview」，在網絡安全方面能力表現突出，能迅速主動發現系統內部深藏的漏洞，但也能變成進行大規模黑客攻擊的武器工具。美國財政部長貝森特與美聯儲主席鮑威爾已緊急召集美國幾家最大銀行的負責人開會，商討相關風險。有分析認為，此舉凸顯AI工具的「雙刃劍」效應。

【大公報訊】當地時間7日，Anthropic發布了其最新旗艦通用模型「Claude Mythos Preview」（以下簡稱Mythos），及一個基於此模型的AI網絡安全項目「玻璃之翼計劃」（Project Glasswing）。

據Anthropic說明，Mythos是該公司最新、最強的模型，尤其在網絡安全任務上的能力，遠遠超過前代模型。在過去的幾周裏，Anthropic內部團隊使用Mythos自主識別了數千個「零日漏洞」（即軟件開發者此前未知的安全缺陷）。這些漏洞廣泛存在於各大主流操作系統、主流網絡瀏覽器以及各類關鍵軟件基礎設施，據稱一些20多年未被發現的漏洞，也被其迅速找出。

Mythos不只能識別漏洞，是一個能「自主發現漏洞並編寫攻擊代碼」的AI模型。它還能在用戶指令下將漏洞「武器化」，寫出可實際攻擊的代碼，並把多個漏洞串聯成一條完整的攻擊鏈，自行完成複雜的黑客任務。據悉，過去黑客發現系統的漏洞可能需要數月，但Mythos能將其壓縮至數分鐘。這意味着，如果該模型落入不法分子手中，其攻擊面幾乎涵蓋全球所有數字基礎設施。

▲Anthropic執行總裁阿莫戴。 路透社

美財政部急召業界開會

Mythos的模型能力引起美國銀行業警戒，對於高度依賴數字系統運作的銀行而言，一旦核心系統被滲透，交易中斷、數據洩露甚至系統性連鎖反應都有可能發生。

彭博社10日報道，美國財政部長貝森特與美聯儲主席鮑威爾7日在華盛頓財政部總部召集華爾街主要銀行CEO，主題是「Anthropic的Mythos與類似模型可能引發的未來風險」。知情人士透露，此次會議旨在確保銀行業充分了解Mythos及同類模型可能帶來的潛在風險，並採取必要的系統防護措施。

知情人士透露，花旗、摩根士丹利、美國銀行、富國銀行，以及高盛的CEO都有出席，幾乎涵蓋美國最大的系統重要性銀行。監管層選擇直接對話CEO級別，而非技術或合規層面，顯示出此警示的緊迫程度。

Anthropic則表示，在Mythos發布前，已就其「攻防網絡能力」與美國政府官員展開持續溝通。

網絡安全「雙刃劍」

監管機構對該模型落入黑客手中可能造成的風險保持謹慎態度，Anthropic為此不向大眾開放Mythos，而是率先啟動「玻璃之翼計劃」項目，邀請僅限於亞馬遜、蘋果、谷歌以及摩根大通等40多家關鍵基建設施合作夥伴參與使用，用來進行防禦性質的安全漏洞修補。

外界認為，新AI模型再次引發警醒，一方面可用於填補漏洞，一方面又可能被黑客利用發動攻擊，甚至具備AI「自主作戰」的雛形。由於AI發現和利用漏洞的速度遠超人類修補速度，可能在一夜之間產生大量可執行攻擊程式，不法分子可輕易發動大規模網絡攻擊、網絡勒索，以及癱瘓銀行、醫院和能源等關鍵基礎設施。前微軟研究與戰略主管蒙迪表示，這一進展或將使頂尖黑客級別的網絡攻擊能力普及化，全球任何單一國家都無法單獨應對。

Anthropic自身存在的安全隱患也讓業界對其能力存疑。此前，Anthropic已多次出現包括代碼洩漏的重大安全事故。值得一提的是，因是否撤除旗下聊天機器人Claude的安全護欄問題，Anthropic已與美國國防部對簿公堂，這也使得監管層與Anthropic之間的關係更加微妙。

（綜合報道）

Anthropic新模型引發擔憂Q&A

Anthropic是家什麼公司？

2021年，擔任美國人工智能OpenAI研究部門副總裁的阿莫戴離職創立Anthropic，創始團隊多為OpenAI的聊天機器人GPT團隊成員。該公司推出Claude系列AI模型，成為OpenAI主要競爭對手之一。

最新模型「Claude Mythos Preview」是什麼？

Anthropic最新推出的AI模型，核心優勢主要集中在超強的漏洞識別與邏輯推理能力，遠超前代模型Claude Opus 4.6。

為何引發擔憂？

新模型能夠以遠超人類的規模識別網絡安全性漏洞，但同時也能自主生成利用這些漏洞的攻擊方法，一旦落入不法分子手中將造成巨大風險。據官方披露，該模型已在主流作業系統、網頁瀏覽器中發現數千個高危漏洞。Anthropic採取審慎發布策略，目前僅向亞馬遜、蘋果和摩根大通等少數科技和金融機構開放。

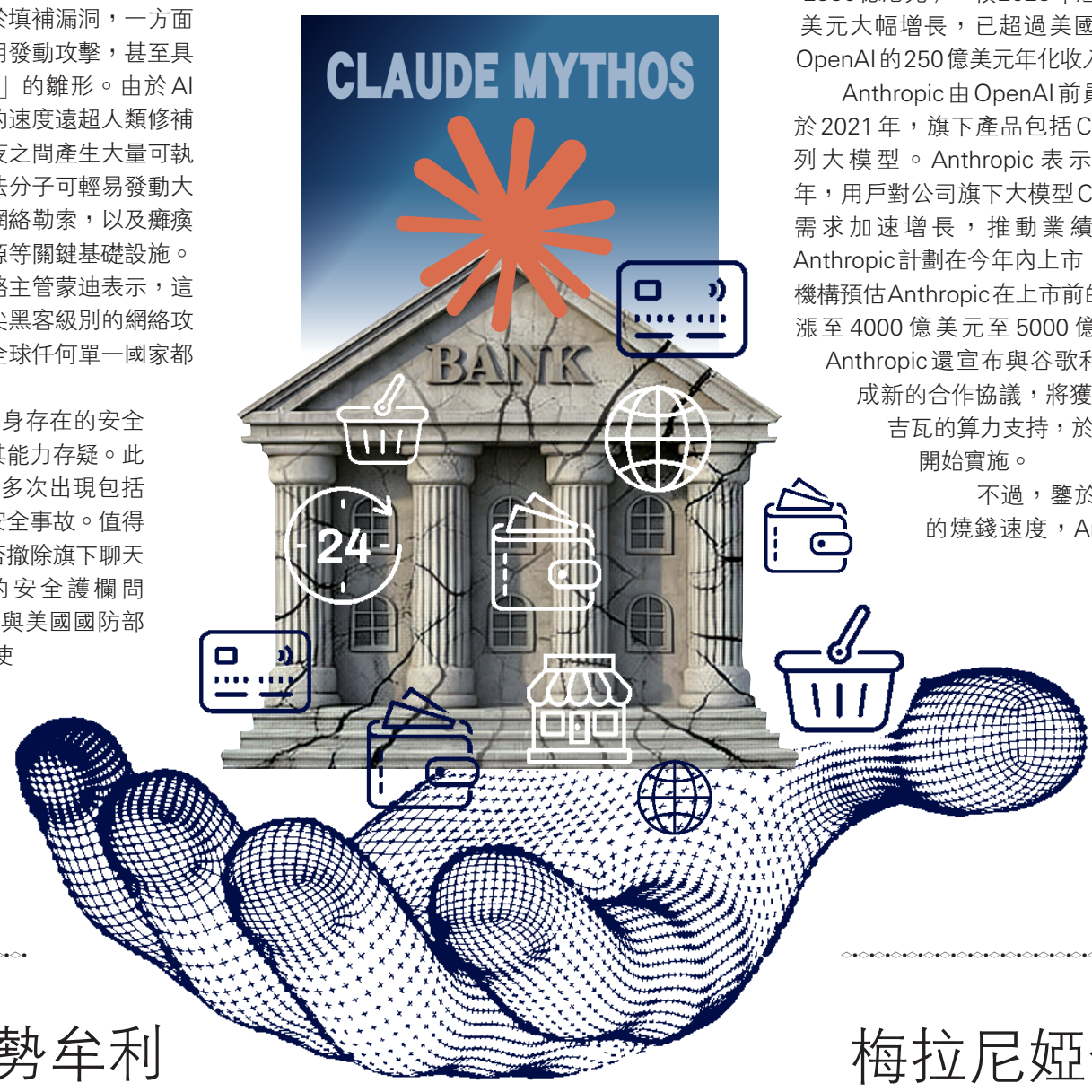
Anthropic近期爭議

數據洩漏

3月31日，Anthropic旗下爆款AI編程工具Claude Code約51.2萬行的源代碼洩漏，被迫向全球開發者「開源」，引發安全性質疑。

被五角大樓列為「供應鏈風險」

美國國防部與Anthropic就美軍是否無限制使用Claude模型問題談不攏，Anthropic被國防部列為「供應鏈風險」企業。Anthropic為此起訴國防部，案件還在審理中。



▲Anthropic推出的AI新模型或影響金融業數據安全。圖為紐約證券交易所。 美聯社

AI競爭愈趨激烈 Anthropic年收入首超OpenAI

【大公報訊】據路透社報道：美國人工智能（AI）初創公司Anthropic於7日公布，公司年化收入（ARR）超過300億美元（約合2350億港元），較2025年底的90億美元大幅增長，已超過美國AI巨頭OpenAI的250億美元年化收入。

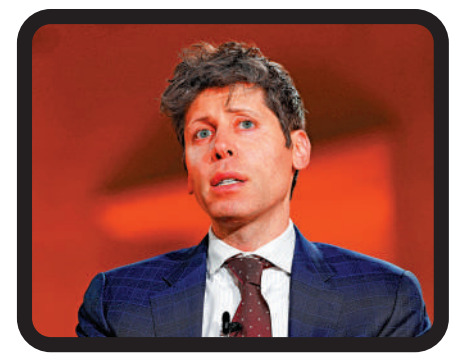
Anthropic由OpenAI前員工成立於2021年，旗下產品包括Claude系列大模型。Anthropic表示，2026年，用戶對公司旗下大模型Claude的需求加速增長，推動業績井噴。Anthropic計劃在今年內上市，華爾街機構預估Anthropic在上市前的估值將漲至4000億美元至5000億美元。

Anthropic還宣布與谷歌和博通達成新的合作協議，將獲得約3.5吉瓦的算力支持，於2027年開始實施。

不過，鑒於AI開發的燒錢速度，Anthropic仍然面臨財務困境。Anthropic的商業模式本質上越來越像「燒錢換算力」的重資產基礎設施生意，核心問題是算力成本失控。目前，推理環節的支出已佔Anthropic總營收的一半以上，與OpenAI面臨的困境類似。

AI巨頭之間的競爭也愈發激烈。Anthropic此前宣布，自4月4日起，Claude訂閱方案不再涵蓋OpenClaw（俗稱「龍蝦」）等第三方工具的使用額度，訂閱用戶若繼續通過OpenClaw使用Claude，必須採用「按量付費」方案，與Claude訂閱分開計費。此前不久，OpenClaw的創始人斯坦伯格據傳加入OpenAI。

另外，Anthropic近期還推出了自己的「龍蝦」，宣布旗下產品Claude Code和Claude Cowork的用戶可以讓Claude控制自己的電腦，打開文件、使用瀏覽器和運行開發工具。



▼OpenAI執行總裁阿爾特曼。 法新社

精準押注伊朗局勢牟利 白宮深陷內幕交易爭議

【大公報訊】綜合路透社、《華爾街日報》報道：美伊戰事期間，金融市場和預測平台出現多筆精準押注，引發對白宮內部可能存在內線交易「發戰爭財」的廣泛質疑。美媒報道，白宮曾向全體員工發送內部警告信，禁止員工濫用職權炒賣。

美國總統特朗普3月22日向伊朗發出48小時「最後通牒」，23日又突然宣布與伊朗舉行了「富有成效」的對話押後攻擊行動5天，引發國際油價劇烈波動，油價一度暴跌15天。就23日特朗普在社交媒體發帖前約15分鐘，期貨市場出現了異常活躍的交易。根據道瓊斯市場數據，在不到兩分鐘內，就有價值超過7.6億美元的原油期貨合約易手。預測市場平台Polymarket也出現類似操作，三個新創建的匿名賬戶精準押注伊朗停火的時間點，共獲利超過60萬美元。這三個賬戶此前也因押注美國將攻擊伊朗而獲利。

據知情人士透露，白宮管理辦公室在3月24日向全體員工發送了一封內部警告郵件，嚴禁工作人員利用職務之便和非公開的政府機密信息，在金融市場和預測市場進行任何形式的投機交易。白宮證實了郵件的真實性，但堅稱目前沒有任何證據表明內部存在信息洩露或內幕交易。

Polymarket近日再爆內線交易爭議。就在特朗普4月7日宣布美伊同意臨時停火的數小時前，至少50個賬戶在該平台大額押注「美伊將在4月7日停火」。

這一事件已引發特朗普的支持者的不滿。一個親特朗普節目的投資者兼聯合主持人埃爾斯沃思表示，「這簡直太噁心了」。

▲Polymarket上「美伊將在4月7日停火」的賭注。 網絡圖片

梅拉尼婭發聲明 否認與愛潑斯坦有關聯

【大公報訊】綜合路透社、美聯社報道：美國第一夫人梅拉尼婭9日突然發表公開聲明，強烈否認自己與已故淫媒富商愛潑斯坦及其同夥麥克斯韋爾有任何關聯，呼籲停止散播關於她與愛潑斯坦關係的「謊言」。

梅拉尼婭當天在白宮宣讀聲明時說，自己與愛潑斯坦並非朋友，從未與愛潑斯坦或麥克斯韋爾有過任何關係。梅拉尼婭強調自己並不是愛潑斯坦的受害者，而愛潑斯坦並沒有把她介紹給特朗普，強調自己與特朗普是1998年在紐約市的一個派對上偶然相識。

2000年，特朗普、梅拉尼婭共同出席愛潑斯坦在場的時裝秀活動，梅拉尼婭稱僅是「社交圈的重疊，並非常態」，形容當時是她第一次認識愛潑斯坦，對後者的犯罪活動「一無所知」。針對社交媒體上一直流傳的有關她和愛潑斯坦的圖片和言論，梅拉尼婭直斥相關指控為「徹頭徹尾的謊言」，是「惡意抹黑」。

美國司法部此前公開的愛潑斯坦案文件，其中包括梅拉尼婭於2002年發送給麥克斯韋爾的一封電郵，內容涉及《紐約》雜誌上關於愛潑斯坦的一篇文章。梅拉尼婭在電郵稱讚這篇文章，並以「愛你的梅拉尼婭（Love, Melania）」署名，引發外界質疑。梅拉尼婭指出，她與麥克斯韋爾過往的互動僅是普通通信。

梅拉尼婭還敦促國會舉行公開聽證會，讓愛潑斯坦案受害者有機會作證，讓更多人了解此案。不過，她宣讀完一早寫好的聲明就轉身離開，沒有回應記者提問。分析認為，梅拉尼婭選擇在此時發聲，令白宮內外感到意外。特朗普政府一直試圖讓愛潑斯坦案降溫，目前尚不清楚梅拉尼婭為何在此時打破沉默。梅拉尼婭的發言人表示，總統知道第一夫人計劃發表聲明。特朗普後來向記者表示，他事先並不知道她打算說什麼。

▲美國第一夫人梅拉尼婭9日在白宮發表聲明。左圖為2000年特朗普、梅拉尼婭與愛潑斯坦、麥克斯韋爾的合照。 網絡圖片