金融業應對 AI 攻擊的挑戰與策略建議

2026/07/13

一、緣起：Mythos 帶來之風險升級

Claude Mythos Preview（以下簡稱 Mythos）係 Anthropic 於 2026 年 4 月對外揭露的前沿模型（frontier model，指能力位於業界最前段、同時可能帶來新型風險與攻防能力的 AI 模型）。各界對 Mythos 之關注，主要來自其辨識漏洞、推演可利用路徑，並規模化產出攻擊或驗證內容之能力。

公開討論中常見的例子包含：在特定測試條件下，Mythos 對瀏覽器與系統漏洞分析、利用鏈推演及可運作 exploit（漏洞利用程式或攻擊利用程式）生成能力，相較前代模型明顯提升；另有案例顯示其具解析深層程式碼脈絡之能力，能挖掘長年潛伏於關鍵系統或元件中的深層漏洞。這些訊號顯示，漏洞發現與武器化的速度、規模與可靠度同步上升，傳統以天或週為單位的修補與應變節奏更容易被追上。

值得注意的是，前沿模型所帶來之風險升級，並不僅在於內容生成能力更強，而在於其已逐步具備規劃、推理、工具呼叫與多步驟執行能力；相較傳統生成式 AI 偏重單次回應，代理型 AI 更可能將風險由資訊層面延伸至系統存取、流程操作與連鎖失效層面。

Anthropic 在揭露 Mythos 時同步啟動 Project Glasswing，使部分關鍵軟體與基礎設施相關單位得以在受控條件下更早發現並修補重大弱點，以降低外溢風險。對金融業而言，即使不直接參與此類計畫，仍可能受到連帶影響，例如關鍵供應商與開源元件的修補節奏更密集，形成 patch flood（短時間大量修補釋出，造成測試與上線資源壅塞）。

因此，Mythos 之關注重點不僅在於單一模型能力提升，更在於其可能壓縮弱點治理與事件應變節奏，並透過供應鏈、修補壅塞及攻擊自動化等途徑，放大金融機構既有資安與營運韌性壓力。

二、Mythos 所引發之主要資安風險

Mythos 所引發之風險，對金融機構而言並非單一技術議題，而係同時牽動弱點管理、事件應變、供應鏈治理與防禦模式調整；尤其當模型已具備一定程度之自主規劃與工具使用能力時，風險已不再侷限於協助產出攻擊內容，而可能進一步表現為自動化探索、反覆測試與修正、動態調整策略與跨步驟攻擊編排，進而加重防禦壓力。其主要風險歸納如下：

- **漏洞揭露與武器化加速**：Mythos 提升了從程式碼理解、弱點辨識到利用鏈推演之效率，使漏洞研究不再高度依賴少數專家之深度手工分析。其直接結果是弱點揭露與可用攻擊手法之產出速度同步上升，防守端可用來評估、排序與修補之時間被明顯壓縮。
- **風險處理時間壓縮與修補壅塞**：一方面，當 time-to-exploit（弱點自揭露、被驗證至遭攻擊者利用之時間）急遽縮短，亦即弱點自揭露、被驗證至遭攻擊者利用之時間，可能由過去以日或週計，壓縮至以小時等級計，組織原本以天或週為單位的弱點管理與變更節奏便容易失效；另一方面，短時間大量修補釋出亦可能形成 patch flood，使分流、驗證與修補作業出現併發壓力，進一步放大回應負擔。
- **攻擊門檻降低與攻擊鏈擴大**：Mythos 可協助產出可運作利用、攻擊腳本與多種變體，降低漏洞研究與利用開發之門檻，並提高將多個小缺陷串接成完整攻擊路徑之可能性。這不僅增加攻擊頻率，也使單點修補更難充分反映整體風險。
- **供應鏈曝險擴大與防禦失衡**：弱點揭露加速後，第三方元件、開源套件、外部服務與代理能力鏈之風險更容易同步外溢；若資產、版本與依賴關係缺乏透明度，便難以迅速判定受影響範圍。另一方面，攻擊端可運用 AI 提升弱點辨識、驗證與攻擊效率，而防禦端若仍主要依賴人工作業，便可能形成明顯的不對稱。

三、金融業因應 Mythos 的挑戰

前述風險進入金融業後，將因金融服務對可用性、交易正確性及信任維持

之要求甚高，加上核心系統高可用性、第三方依賴程度高、監理責任明確及跨單位決策流程較複雜，而進一步放大為治理與營運壓力。於高頻、短時窗攻擊情境下，金融機構較易面臨處置速度、授權機制與營運承受能力之落差：

- **核心系統高穩定性要求下的變更受限**：金融業許多核心交易、清算、支付與對客系統難以頻繁停機或快速改版，修補作業往往必須配合回歸測試、跨系統驗證、上線窗口與回退準備同步進行。當外部修補波次變密集時，真正的壓力不只是修補數量增加，而是原本為維持穩定性所設計的變更節奏，容易在短時間內被擠壓，形成人力與測試資源不足、維運排程與業務容忍度衝突的情形。
- **短時窗攻擊情境下之授權與服務維持壓力**：金融機構面對事件時，不僅要判斷是否隔離、封鎖或降級，還須同步考量交易中斷、客戶影響、通報義務與對外溝通。當攻擊利用時窗急遽縮短，可供研判、決策與處置之時間大幅壓縮，宜事先建立預授權門檻、例外流程與降級劇本，避免因流程過長而錯失止血時機。
- **既有風險衡量與優先序機制易失真**：原風險評估慣用之弱點分級、修補期限與重大性判斷，多建立在相對穩定的事件節奏與修補窗口之上；但在高頻、多波次且跨系統併發的情境下，原有指標可能無法真實反映實際曝險與處置壓力，致使管理階層低估資源配置與營運韌性之累積衝擊。
- **資產與版本基礎資料不足時難以精準應變**：真正影響應變速度的，通常不是有沒有收到弱點資訊，而是能否在短時間內對應到實際上線資產、版本資訊與依賴關係。若基礎資料長期未能維持一致，則修補優先序、隔離範圍與替代流程都難以精準判定，迫使組織在資訊不完整下作出高成本決策。
- **治理責任與問責要求並存**：面對新型資安風險時，除需控制技術衝擊，也需兼顧內控制度、重大事件通報、客戶權益維護及高階管理階層之治理責任。此使其應變作為不僅要求迅速，亦須符合程序要求與責任歸屬。

四、國際金融監理反應

綜觀近期國際金融監理動向，各國雖因監理架構、金融市場特性及推動進度不同，具體作法未盡一致，惟整體上均已將 Mythos 等前沿模型視為可能改變

金融資安風險節奏與營運韌性要求的重要變數。監理關注亦由模型技術能力本身，進一步延伸至弱點揭露速度、攻擊利用時窗急遽縮短、供應鏈曝險、攻擊自動化及重大事件應變壓力之綜合影響。

就監理取向而言，國際上可觀察到由原則性提醒逐步轉向較具體治理參考之趨勢。美國除由高層政策與監理部門召集大型銀行管理階層交換意見[1]外，亦出現鼓勵銀行在受控條件下實際測試 Mythos 影響的作法[2]，並將其提升為系統性資安議題。英國則先由中央銀行、財政部、金融行為監理機關及國家網路安全部門透過工作小組機制納入議程[3]，其後再以正式聯合聲明提出治理、弱點管理、供應鏈控制、偵測回應與持續營運整備方向[4]。日本則由跨機關會議、任務編組與官民協作[5]，進一步發展為由金融廳與日本銀行聯名提出可優先採行之短期整備措施[6]，強調金融機構應及早就關鍵服務、系統曝險及營運韌性進行盤點與準備。

除前述代表市場外，其他市場之監理思維亦逐步由機構內部控制延伸至供應鏈可視性、第三方依賴管理與跨機構協作機制。新加坡、香港、澳洲及韓國等市場[7]，雖多先透過媒體揭露監理動作，但其內容已顯示主管機關除關切前沿模型可能加速弱點發現與利用外，亦傾向鼓勵金融機構加強主動盤點、改善資安基礎防護、預作韌性測試及建立公私協作安排。

整體而言，國際金融監理對前沿模型所衍生資安風險之關注，已由原則性提醒逐步轉向較具體之治理參考與實務整備方向。近期 Five Eyes 資安機關聯合聲明[8]亦指出，AI 已非未來議題，而是正在加速攻擊速度、規模與複雜度之現實風險；其並強調資安韌性已是業務持續、金融市場信心與長期價值之核心基礎，董事會及高階管理階層應將資安視為核心營運風險與領導責任，而非單純技術議題。該聲明所提出之方向，包括瞭解並評估風險、準備度與問責機制，優先落實基礎資安控制，賦予資安主管必要權限與資源，並隨威脅與指引演進

[1] https://www.cnbc.com/2026/04/10/powell-bessent-us-bank-ceos-anthropic-mythos-ai-cyber.html
[2] https://www.bloomberg.com/news/articles/2026-04-10/wall-street-banks-try-out-anthropic-s-mythos-as-us-urges-testing
[3] https://www.bloomberg.com/news/articles/2026-04-11/bank-of-england-set-to-discuss-anthropic-s-mythos-with-banks
[4] https://www.fca.org.uk/news/statements/fca-boe-treasury-joint-statement-frontier-ai-models-cyber-resilience
[5] https://www.bloomberg.com/news/articles/2026-04-22/japan-finance-minister-to-meet-banks-to-discuss-mythos-threat
[6] https://www.fsa.go.jp/news/r7/sonota/20260522-5/20260522.html
[7] https://finance.yahoo.com/economy/policy/articles/regulators-monitor-anthropics-mythos-banking-075513004.html
[8] https://www.cyber.gov.au/about-us/view-all-content/news/five-eyes-cyber-security-agencies-statement

持續投入，與金融業面對前沿模型風險時所需之治理取向高度一致。基此，我國金融業可在既有資安政策與韌性要求基礎上，參酌前述國際監理觀察，進一步轉化為可分階段推動之具體整備與能力建設方向。

五、金融業因應策略建議

面對前述風險變化，金融機構宜在既有資安政策與韌性要求基礎上，依風險程度與系統重要性強化因應作為。整體策略可分為三個面向：一是加速落實既有政策，鞏固共同治理與基礎控制；二是推動短期整備，提升弱點揭露加速與攻擊利用時窗急遽縮短情境下之快速研判與應處能力；三是前瞻規劃防禦性AI與相關治理機制，以支撐中長期資安韌性。

在前沿模型使弱點揭露後遭快速利用之情境下，因應重點不宜僅限於加快修補，而應同步強化資產與依賴可視性、風險分流、替代控管、委外協調及營運韌性。相關措施宜以關鍵服務、對外曝露系統及高風險委外項目為優先，並逐步納入常態化治理與演練機制。

（一）既有政策加速落實與深化執行

金管會前已發布「金融資安韌性發展藍圖」，其核心方向包括零信任、持續監控、聯防協作、資安左移、第三方治理及韌性驗證等。面對前沿模型加速弱點揭露與攻擊利用之新情勢，金融機構可優先從藍圖既有方向加速落實並深化執行：

- **落實零信任，降低單點失守衝擊：**在難以預先保證系統完全無漏洞之前提下，可透過抗釣魚多因素驗證、設備健康掌握、網路分段、連線控管與持續驗證，提升關鍵資源之防禦縱深，限制橫向移動、權限擴張與降低資料外洩風險。對外部或客戶端服務，宜將前端保護與傳輸安全視為基礎防護，並同步強化後端服務與 API 之身分驗證、授權檢核、異常偵測及流量控管，以降低自動化濫用與未授權存取風險。【參照「金融業導入零信任架構參考指引」- 等級**I**、**II**】
- **強化持續監控，提升研判與即時應處效能：**面對攻擊利用時窗急遽縮短，

資安監控宜由告警可見進一步提升為可早期發現異常、關聯弱點與攻擊線索、支援快速研判並驅動應處之能力，並持續優化偵測、分流與止血效能。【參照「金融業導入零信任架構參考指引」- 等級**III、IV**】

- **深化資安左移，將資安要求前移至開發、測試與部署流程：**將資安要求逐步前移至需求、開發、測試與部署各階段，並同步強化 CI/CD（持續整合與持續交付／部署）安全控制、自動化檢測、簽章驗證、回歸測試及回退安排，作為提升弱點治理效率與修補韌性之基礎。【參照「金融資安韌性發展藍圖」- **資安左移：鼓勵導入軟體安全開發、測試及部署流程(CI/CD)** 】
- **強化第三方治理，提升供應鏈可視性與弱點追蹤能力：**將 SBOM（軟體物料清單，用以記錄軟體元件、版本及相依關係）與依賴管理納入常態作業，作為快速定位受影響範圍、排序修補優先序及支撐漏洞應變之基礎，並逐步明確關鍵供應商之修補 SLA（服務水準協議）、回報時限、緩解措施及替代機制。【參照「金融資安韌性發展藍圖」- **資安左移：建立軟體供應鏈透明化(SBOM)與弱點追蹤機制】**
- **深化聯防協作，強化重大事件通報與情資共享：**在既有 F-ISAC（金融資安資訊分享與分析中心）情資分享基礎上，針對高風險弱點、重大供應鏈事件及密集修補波次，逐步明確分級報送門檻與通報時限，並釐清與委外廠商、關鍵供應商及聯防單位間之責任分工，以利形成共同情勢感知與優先應處順序。【參照「金融資安韌性發展藍圖」- **加強資安情資分析與協同防禦】**

（二）短期整備措施

短期整備措施著重於在前沿模型使攻擊利用時窗急遽縮短之情境下，降低當前曝險、縮短研判與協調時間，並提升重大事件下之基本處置與復原能力。呼應 Five Eyes 資安機關聯合聲明所強調之「落實基礎資安控制、迅速採取行動，並將資安納入核心營運策略」原則，金融機構宜優先聚焦對關鍵服務具直接助益之事項，包括盤點、分流、替代控管、委外協調及韌性演練：

- **明確短期應處之統籌與督導機制：**宜由資訊、資安、營運、法遵及風險管理等單位共同組成短期應處機制，並由高階管理層督導，以利整合風險盤

點、資源調度與跨單位決策，避免因分工不清或授權遲滯影響整備效率。

- **先識別優先服務、關鍵系統與外部曝露面：**可先以風險為基礎，識別對外服務、核心交易節點、身分驗證、遠端管理、高權限帳號路徑及其他關鍵服務所依賴之系統，作為優先整備對象；對共同營運或委外提供之系統，亦宜同步釐清責任分工。
- **優先提升資產可視性並清理既有系統問題：**對已識別之優先服務與系統，先確認軟體版本、網路配置、外部介接與依賴元件，並加速處理不必要之外部曝露、閒置高權限帳號、已知未修補項目及已停止支援版本，以提升後續修補與緩解效率。
- **盤點修補量能與短期資源排序機制：**可檢視內部人力、委外維護量能及夜間、假日支援安排，確認在短時間內出現大量弱點與修補需求時，是否足以支撐修補、驗證與部署作業；對委外維運、雲端服務及其他第三方服務，亦宜同步釐清緊急支援、責任分工、回報時限及替代處置安排，並校準短期風險排序邏輯，以維持高壓期處置品質。
- **採行風險導向之弱點分流、修補與測試：**可依服務關鍵性、外部可利用可能性、實際曝險程度及影響範圍，建立短期優先處置清單，而非僅依單一評分指標排序；對修補前測試，亦宜在控制穩定性風險之前提下調整範圍與時程，以兼顧延遲修補與測試不足之風險。
- **對暫難立即修補項目強化替代控管：**對暫時無法立即完成修補之系統，可同步採取隔離、權限收斂、流量限制、停用非必要服務、臨時阻擋規則、端點偵測與回應或網路分段等措施，並視需要辦理殘餘風險評估與接受程序，以降低曝險期間之實際風險。相關替代控管宜納入事前授權、啟動條件、影響評估及回復安排，以避免緊急處置對服務造成不必要之影響。
- **預作停機、降級運作與對外說明準備：**可就關鍵服務或系統預先規劃暫停、降級運作、替代流程啟動及對客戶與利害關係人之說明安排，並明確相關觸發條件、授權門檻與應變流程，以提升重大風險情境下之決策速度、一致性及持續營運能力。
- **預作韌性演練與短期驗證：**可將攻擊利用時窗急遽縮短、patch flood、多起高風險事件同時發生及供應鏈連鎖影響等情境，納入近期可執行之演練、桌上推演或測試，並優先驗證隔離啟動時間、降級運作能力、關鍵服務持續性、備份備援可用性及替代流程可行性；其中，備份機制應涵蓋離

線備份或不可變備份等安排，以確認機構在高風險弱點迅速遭利用情境下之基本應處與復原能力。

短期整備期間，宜建立可追蹤之檢視機制，以掌握高風險弱點判定、重大修補處理、通報協調、隔離或降級啟動等關鍵環節之執行情形，並作為滾動調整資源配置與應變流程之依據。

（三）前瞻規劃防禦性 AI

防禦性 AI 之前瞻規劃，旨在將短期整備所建立之盤點、分流、監控與應變基礎，進一步轉化為可持續、可擴充之防禦能力。Five Eyes 資安機關聯合聲明指出，攻擊者已運用 AI 提升攻擊速度、規模與複雜度；防禦端亦宜在治理可控之前提下，審慎運用 AI 強化弱點辨識、軟體安全、異常監控、告警關聯與事件應處等能力。

防禦性 AI 之導入不宜僅視為工具更新，尚須與既有資安控制、治理流程、資料保護、人工作業覆核及營運韌性安排相互銜接，以避免因自動化導入而形成新的機密外洩、誤判或控制失衡風險。金融機構宜據此將防禦性 AI 納入中長期能力建設與治理規劃，並依風險程度與控制成熟度，循序評估適用場景及導入條件，例如：

- **增進軟體安全開發之可視性與作業效率**：可運用相關工具協助檢視原始碼、腳本、設定檔與相依元件，提早辨識常見弱點、不安全設定與版本風險，並輔助產出修補建議、受影響範圍盤點及回歸測試重點，以縮短由發現到處置之時間。
- **建立可持續之漏洞修補與弱點治理產能**：可整合發現、驗證、排序、修補、部署、回退與成效衡量流程，逐步提升整體修補吞吐量與協同效率，並於核心系統與高穩定性場域建立分級測試、例外核准及快速回退安排，以兼顧風險控制與服務可用性。
- **建立可規模化之監控分析與事件應處能力**：協助告警彙整、關聯分析、事件脈絡整理、取證資料蒐整與報告初稿產製，提升高頻事件下之分流效率；惟相關應用仍宜保留人工覆核、軌跡保存、停用與回退機制，並逐步

納入代理型 AI 行為特徵之辨識，以降低誤判與自動化失控風險。

六、結語

綜上，Mythos 等前沿模型已改變金融資安風險之速度、規模與不確定性，使金融機構面對之挑戰不再僅是技術防護問題，而是攸關治理責任、營運持續與市場信任之核心風險。面對弱點揭露加速、攻擊利用時窗急遽縮短與供應鏈曝險擴大等趨勢，金融機構不宜等待風險完全具體化後再行因應，而應及早由高階管理層主導，將相關整備納入營運韌性與資安治理之優先事項。

後續推動上，金融機構應以既有資安政策與韌性要求之深化落實為基礎，以短期整備降低當前曝險並提升事件應處與復原能力，並以防禦性 AI 治理及韌性驗證作為中長期能力建設方向；同時加速強化資產與依賴可視性、弱點分流與快速應處、供應鏈管理、身分與存取控制、事件演練、備份備援及韌性驗證等關鍵能力。對防禦性 AI 之應用，亦應在治理可控、風險可管及責任明確之前提下審慎導入。

由於前沿模型能力演進快速，既有風險假設可能在短時間內失準，金融機構須持續關注國際監理動向、威脅情資及防禦技術發展，並滾動檢視內部風險評估、控制措施與整備優先序，確保資安韌性真正融入核心營運策略。金管會亦將持續掌握國際監理實務及產業整備情形，必要時調整監理重點、參考指引或公私協作機制，協助金融業即時回應快速演進之 AI 資安風險，維持金融服務之安全、穩定與市場信任。