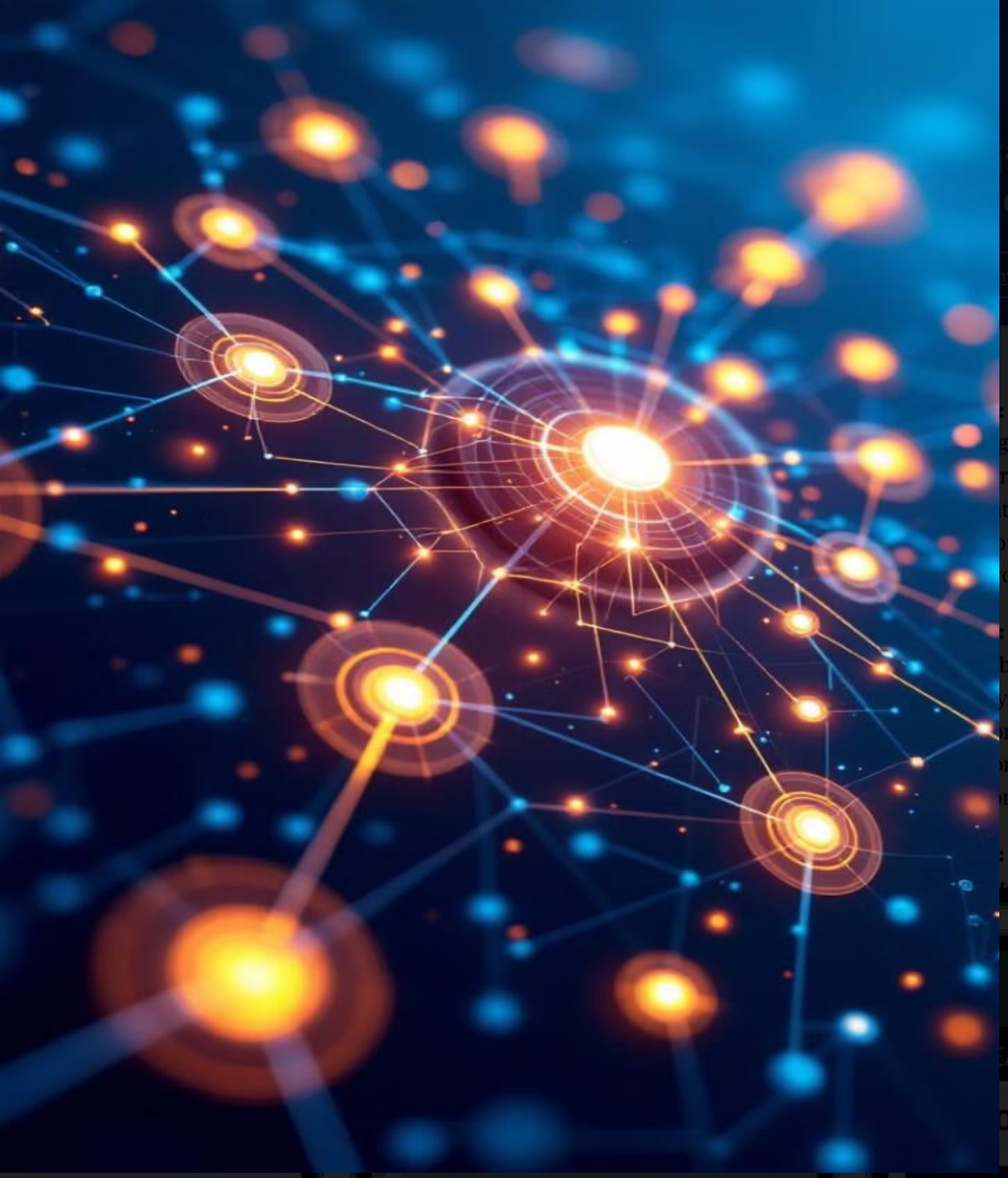
OPSWAT.

ARTIFICIAL INTELLIGENCE

AI 世代下的資訊安全挑戰

次世代下的資訊安全挑戰

Kent Chou
OPSWAT, Solution Engineer, North Asia
Kent.chou@opswat.com



人工智能正徹底革新
現代工作方式。
2025年全球AI市場
預計達$1,847.58
億美元。

Key AI Milestones

2018 — BERT 模型推出

- **技術突破**：雙向語言理解（Transformer 架構）
- **意義**：NLP 開始邁入理解語意的時代

2020 — GPT-3 推出（OpenAI）

- **AI 具備自然對話、寫作、程式設計等能力**

2022 — ChatGPT 開放測試

- **快速全球爆紅，廣泛應用於教育、創作、客服等場景**

2023–2024 — 多模態 AI 與 Agent 系統興起

- 能理解圖片、語音、影片，具備推理與工具使用能力
- 邁向通用人工智慧（AGI）之路

AI 技術應用及分類

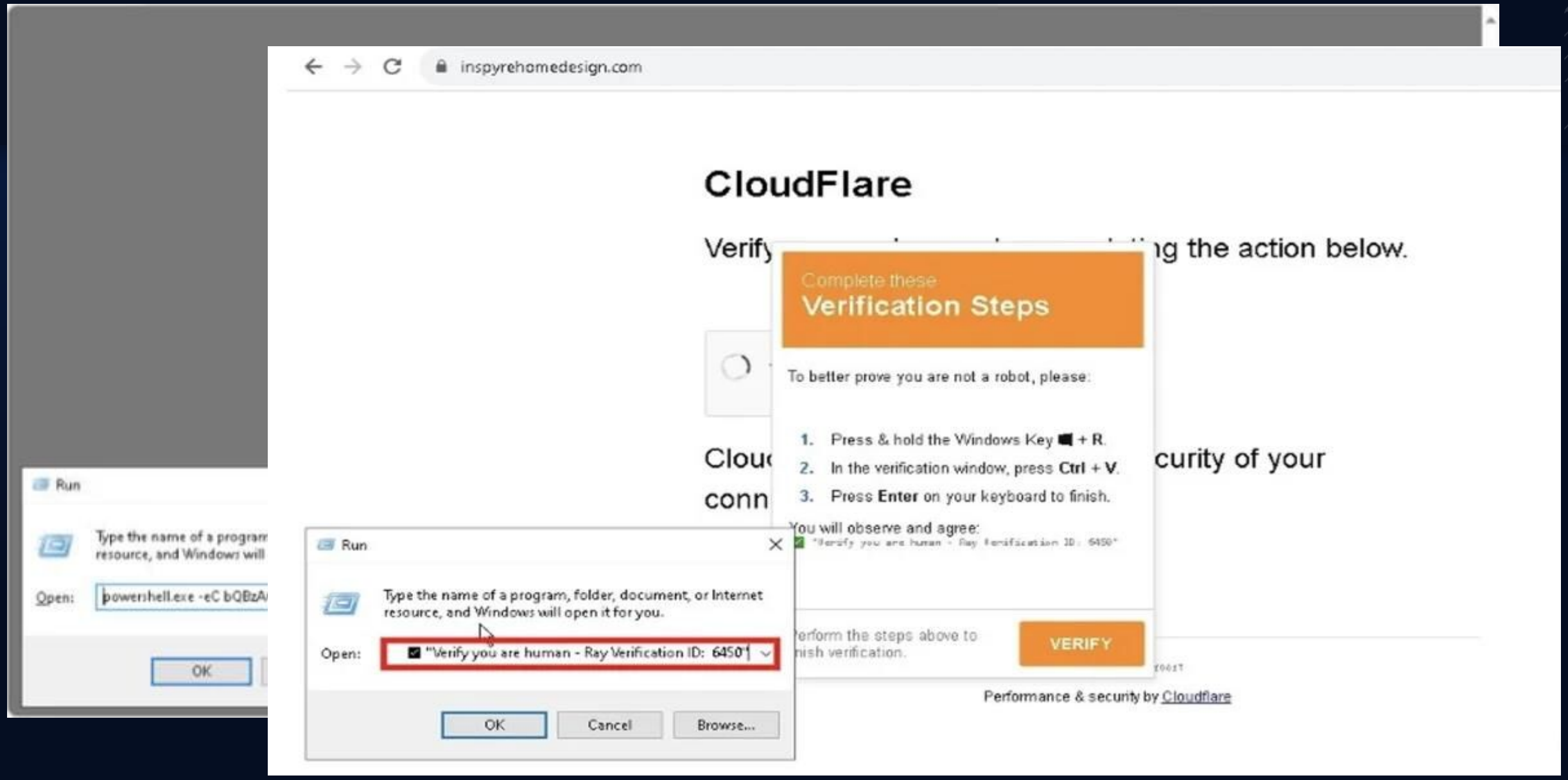
技術	應用	成熟代表
文字生成	寫作、摘要、客服、教學、程式碼	ChatGPT（OpenAI）、Gemini（Google）、Claude（Anthropic）
圖像生成	設計、廣告、插畫、視覺創意	DALL·E 3、Midjourney、Stable Diffusion
語音合成		
影片生成		
程式碼生成		

技術	應用	成熟代表
語音辨識	自動字幕、語音助理	Whisper（OpenAI）、Google ASR、iFlytek（科大訊飛）
圖像辨識	醫療影像分析、監控、人臉辨識	YOLOv7、Vision Transformer、Face++
物件追蹤/分類	自駕車、智慧工廠	Tesla FSD、NVIDIA Isaac
自然語言理解（NLU）	問答系統、客服機器人、文件分析	BERT、ERNIE、RoBERTa

近年來的真實案例

案例名稱	時間	地點	技術手法	結果與影響
ChatGPT 生成惡意程式碼	2023	全球	提示繞過生成木馬與勒索病毒	模型遭濫用，引發OpenAI政策調整
Deepfake CEO詐騙	2019	英國	AI 音頻偽造	成功騙取 €240,000資金
破解CAPTCHA	2021~今	多國研究機構	AI 模型訓練破解圖像驗證碼	傳統驗證方式安全性受威脅
AI生成釣魚郵件	2023	美國	模擬高層語氣與內容風格	高點擊率與成功率，資安訓練急需強化
模擬使用者滲透API	近年研究	多個平台	模仿正常使用行為測試漏洞	發現資料洩漏點，觸發平台修補

驗證也成了一種攻擊來源



https://reliaquest.com/blog/using-captcha-for-compromise/

除此之外，AI模型也具有資安風險

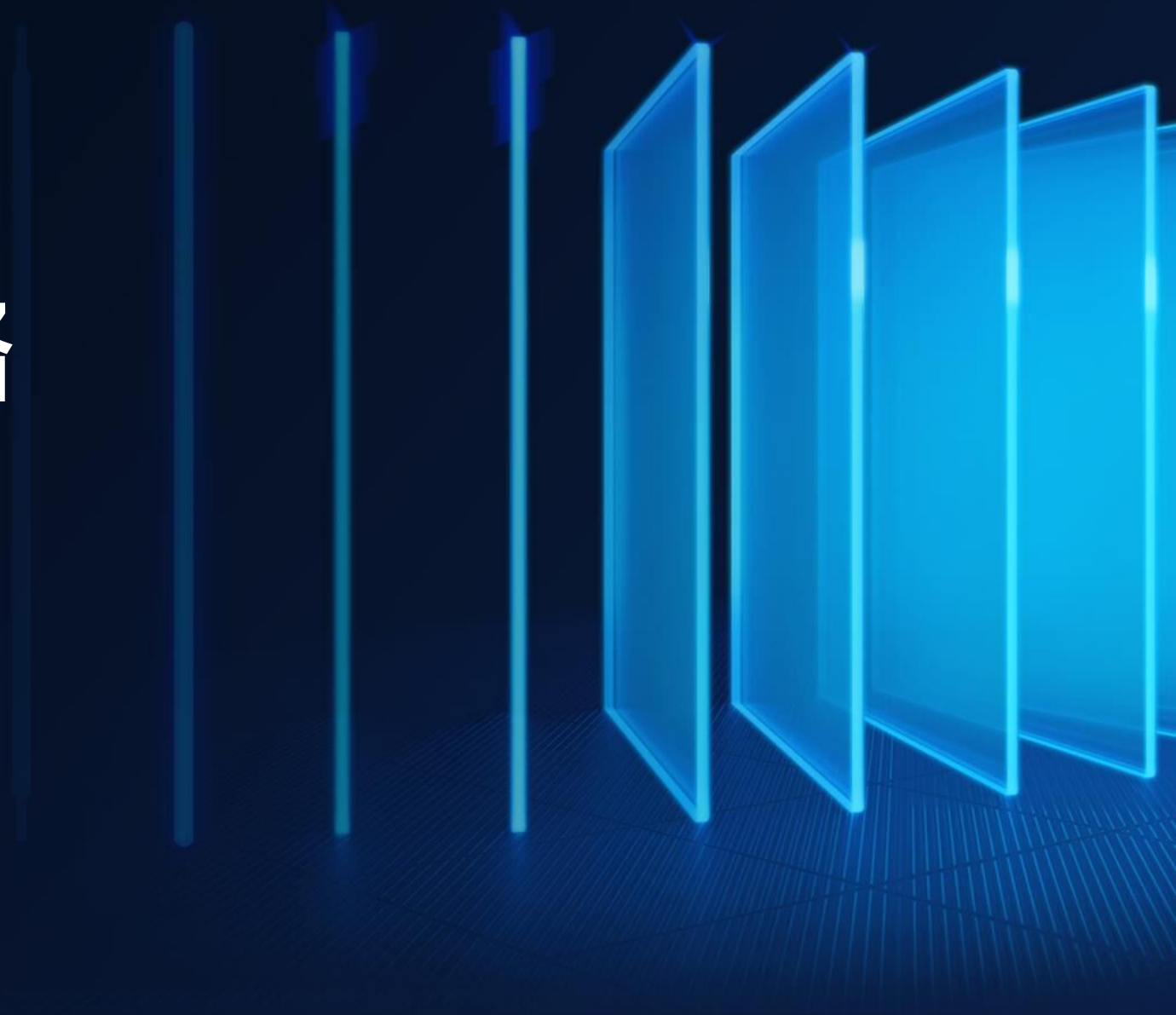
有好的工具，也需要足夠的知識去保護我們的環境

風險類型	說明	參考資料
Prompt Injection	操控模型輸出惡意或敏感資訊	LLM01:2025 Prompt Injection
Model Leaking	模型回應中洩漏訓練資料或內部內容	LLM07:2025 System Prompt Leakage
Model Theft	模型被盜或逆向重建	ML05:2023 Model Theft LLM10: Model Theft
Adversarial Attack	特殊輸入讓系統誤判、繞過安全偵測	ML01:2023 Input Manipulation Attack Threats through use

OPSWAT.

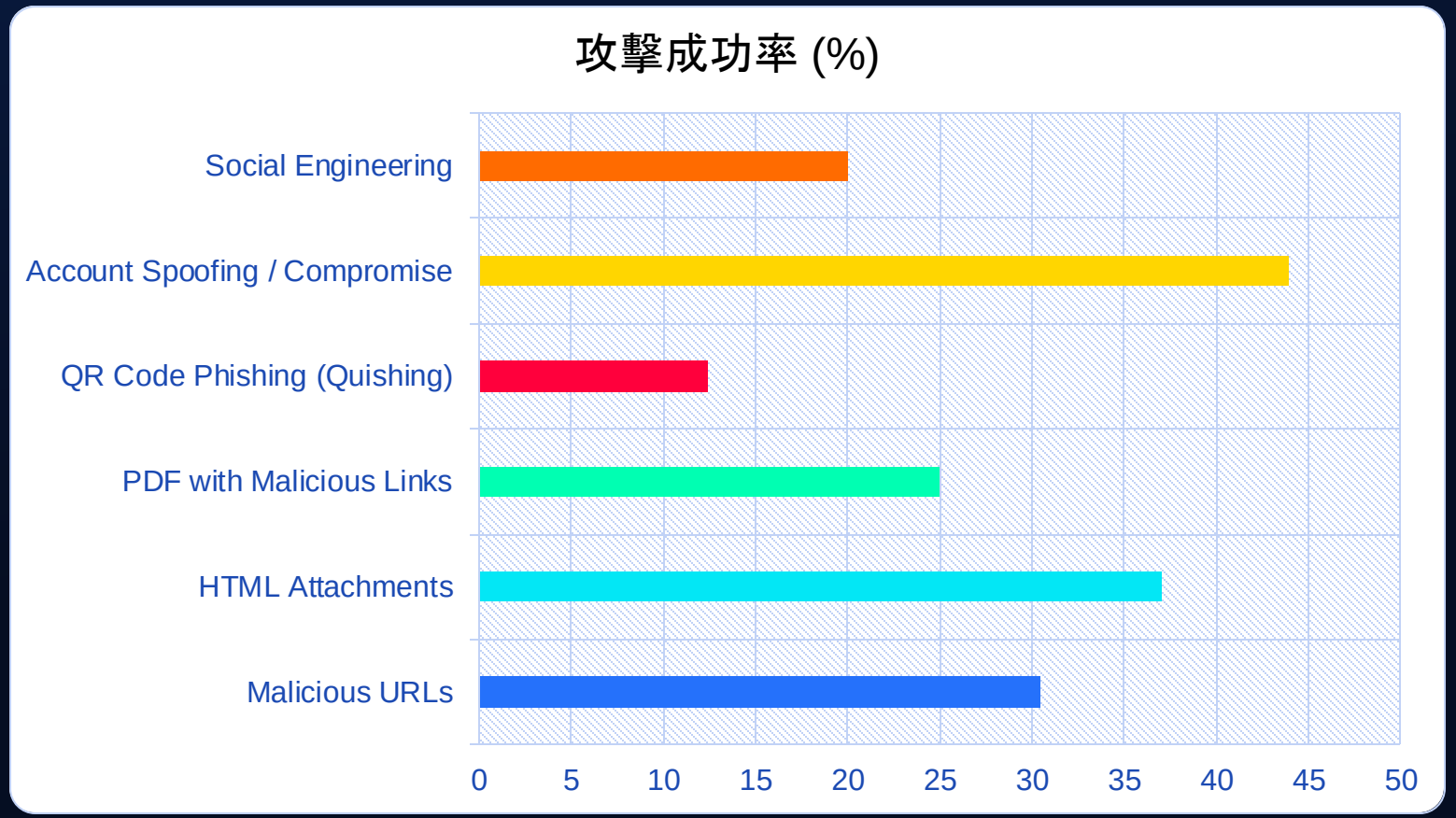
AI 的下資安思路

落實資安停看聽！



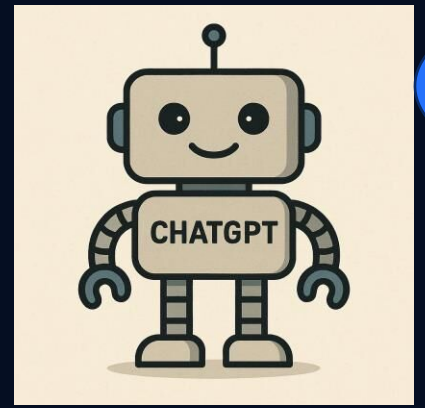
無法利用傳統防護達成的目標

Common Phishing Email Techniques



AI 的幫助下，化繁為簡的矛盾之處

- 網址
- 檔案
- 圖片
- QR Code
- 深度偽造技術
- 社交工程的延伸



I can help you

資料來源: https://www.redhotcyber.com/en/post/crazyhunter-the-ransomware-with-the-three-dimensional-data-annihilation-system-that-redefines-data-destruction

CrazyHunter

Crazyhunter Team

Home About me Contact us

1. Ultra-fast attack approach: "72-hour Vulnerability Response Vacuum"

Crazyhunter claims to **breach victim security in less than 72 hours**, leveraging:

- **Exclusive exploit chains** with a **survival duration 300% longer than MITRE's average estimates**.
- **Advanced bypass techniques** for leading endpoint protection systems, including:
  - *CrowdStrike*
  - *SentinelOne*
  - *Microsoft Defender XDR*
  - *Symantec EDR*
  - *Trend Micro XDR*

This suggests that the group exploits **zero-day** or **well-selected N-day vulnerabilities**, in addition to **advanced evasion tactics**, which may include **polymorphic malware** and **fileless attack techniques**.

專屬漏洞利用鏈，其存活時間比 MITRE 預估平均值長 300%

可繞過主流端點防護系統的進階技術手法

Crazyhunter Team

Home About me Contact us

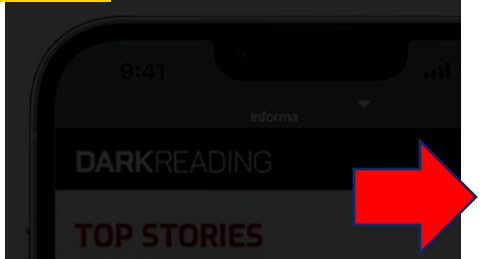
Victim List

Conclusions: Why Crazyhunter Is a Threat to Watch

Crazyhunter is **not just another ransomware group**. Unlike other operations that focus solely on file encryption, this group introduces **additional pressure tactics**, including:

- **Irreversible data destruction, beyond encryption.**
- **AI-powered deepfake evidence creation to compromise executives.**
- **Blockchain-**
criminal market.

運用 AI 生成深偽內容，偽造企業高層進行構陷或勒索



案例分析：針對近期醫療院所面對 CrazyHunter 勒索軟體攻擊事件技術過程探討說明

https://teamt5.org/tw/posts/the-case-study-hospital-crazyhunter-ransomware-attack/

檔案名稱	描述	SHA-256 雜湊值	參考資料
a.Ashx	Godzilla Web Shell	6bb811e2fbb498f466980a176caefbfb	teamt5.org
tunnel.aspx	reGeorg Web Tunnel	5cc2523816a184fed135f0119756c337	teamt5.org
bb.exe	Shellcode 載入器，負責載入 crazyhunter.sys	2cc975fdb21f6dd20775aa52c7b3db6866c50761e22338b08ffc7f7748b2acaa	labs.withsecure.com
crazyhunter.exe	使用 Prince Ransomware 建構器生成的勒索軟體加密器	53160607452717 23c9934047155dae95a3920cb6343ca08c93531e1c235861ba	labs.withsecure.com
crazyhunter.sys	基於 crazyhunter.exe 的 Shellcode 二進位檔案	906e89f6eb39919c6d12a660b68ae81f	teamt5.org
go.exe	防禦規避工具	ca257aaa1ded22ca22086b9e95cb456d	teamt5.org
go3.exe	與 crazyhunter.exe 相同的勒索軟體加密器	da1a93627cec6665ae28baaf23ff27c5	teamt5.org

防毒軟體的規避

接下來我們來看一下批次檔中看到的一些執行檔。

MD5	File Name	Time Stamp
7f05a928c77cb87ffb510168c1b0b11b	aa.exe, cc.exe	2025-01-27 17:36:21 UTC
e12c5be075c23d1c5f398e46e2ee5d40	av-1m.exe	2025-02-08 11:01:11 UTC
87b3db166041c61f3a033cf3c94e89c6	av-500kb.exe	2025-02-08 10:49:56 UTC
ca257aaa1ded22ca22086b9e95cb456d	go.exe	1970-01-01 00:00:00 UTC
da1a93627cec6665ae28baaf23ff27c5	go2.exe	1970-01-01 00:00:00 UTC

使用了一點的 **OLLVM** 技巧，不過字串還是都沒編碼。這些字串為 **防毒的 Process 名稱**，將它取出
現它的目的是 **針對以下兩家防毒產品進行終結**，有看到的字串為下列執行檔名稱：

- **Trend Micro OfficeScan 防毒**
  - EndpointBasecamp.exe、NisSrv.exe、PccNt.exe、PccNTMon.exe、TmListen.exe、Ntrtscan.exe、NTRTScan.exe、TMBMSRV.exe、TmCCSF.exe、CNTAoSMgr.exe
- **Microsoft Defender for Endpoint（MDE）防毒**
  - MsMpEng.exe、MsSense.exe、SecurityHealthService.exe、SenseTVM.exe

199055cfba8dfb0288c3060f115087db5072298a

Threat name: Trojan/Zamtamper!5tZB3lvU — Cast your vote on this file: 0 0

Export PDF

79c3fd97d33e114f8681c565f983cd8b8f9d8d93

Threat name: Trojan/Zapchast!vhTnLqup — Cast your vote on this file: 0 0

Export PDF

Metascan Multiscan — Result — Engine — Last Update

Threats detected

b6737248f7baed88177658598002df5433155450

Threat name: Trojan/Zapchast!4c9GKy32 — Cast your vote on this file: 0 0

Export PDF

Metascan Multiscan

Threats detected

05 /21 ENGINES

Multiscanning, is an advanced threat detection and prevention technology that increases detection rates, decreases outbreak detection times and provides resiliency to anti-malware vendor issues.

Result	Engine	Last Update
✕ Trojan/Win.Generic	AhnLab	Feb 11, 2025
✕ TR/Agent.Hevki	Avira	Feb 10, 2025
✕ W32.Malware.172C240B	Bkav Pro	Feb 10, 2025
✕ Trojan_Win64_Zapchast_of	CMC	Feb 11, 2025
✕ Trojan.WinGo.Rozena	IKARUS	Feb 10, 2025
✓ No Threat Detected	Bitdefender	Feb 10, 2025

在每個交換檔案的邊界，提供最有效的防護

檔案安全

超過30個防毒軟體的選項

AhnLab · Avira · Bitdefender · Bkav · BlackBerry Cylance · COMODO · CrowdStrike · CS · EMSISOFT · eset · Filseclab · 火绒安全 · IKARUS · K7 SECURITY · LIONIC · McAfee · Microsoft Defender · NANO · nG · Quick Heal · RocketCyber · SOPHOS · TACHYON · varist · WEBROOT

180+ supported file types

內容清洗和重建

DOC · DOCX · XLS · XLSX · PPT · RFT · PDF · HTML · XML · JPG · PNG · GIF · SVG · MP4

基於仿真的即時沙盒

10x 比傳統沙箱更快

100x 比其他 Sandbox 資源效率更高

25k+ 一台設備每天處理檔案數

敏感訊息防護

110+ supported file types

OCR Optical Character Recognition

AI named-entity recognition

在安裝之前檢測應用程式漏洞

檢測源代碼及 Docker Image

實時威脅數據和洞察

原產國和供應商身份識別

MetaScan®

Deep CDR

Adaptive Sandbox

Proactive DLP

File-based Vulnerability Assessment

Threat Intelligence

SBOM

Country of Origin

MetaDefender™
CORE TECHNOLOGIES

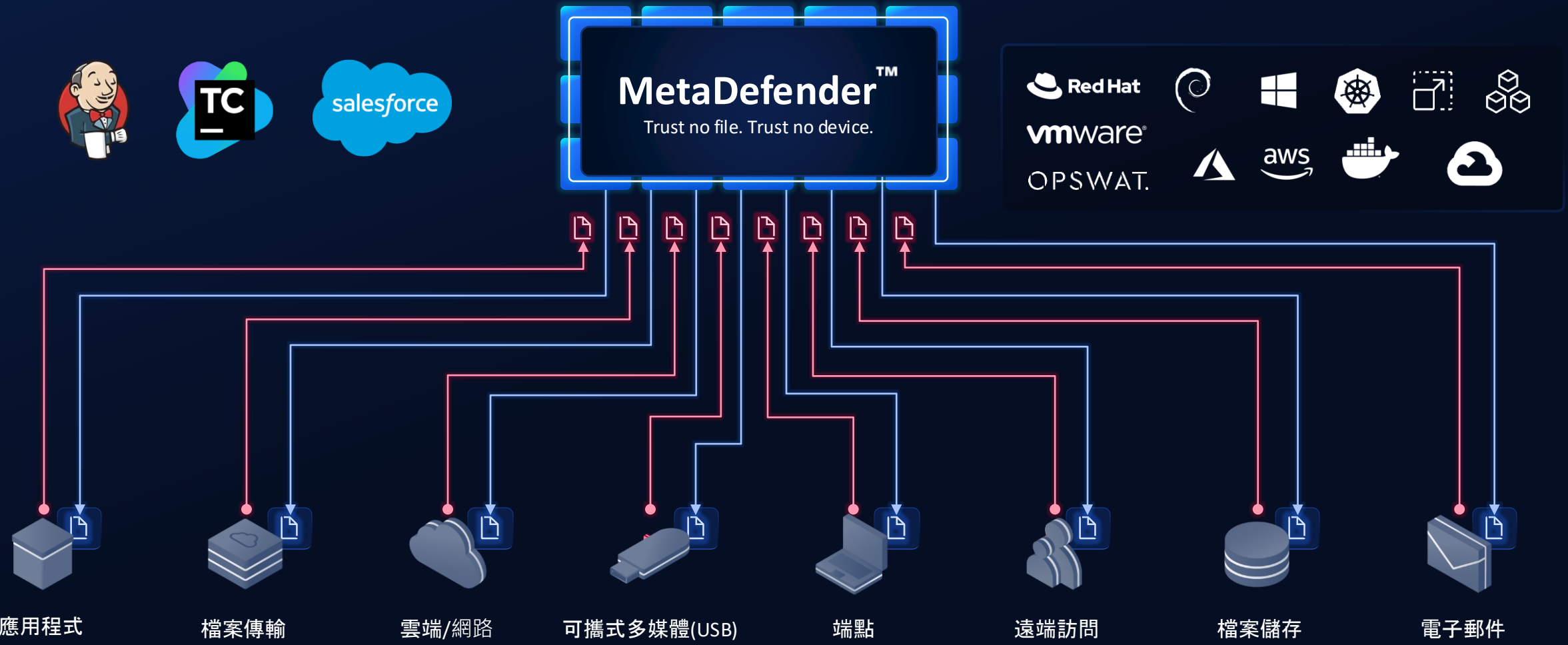
YARA

External Scanner

Post Action

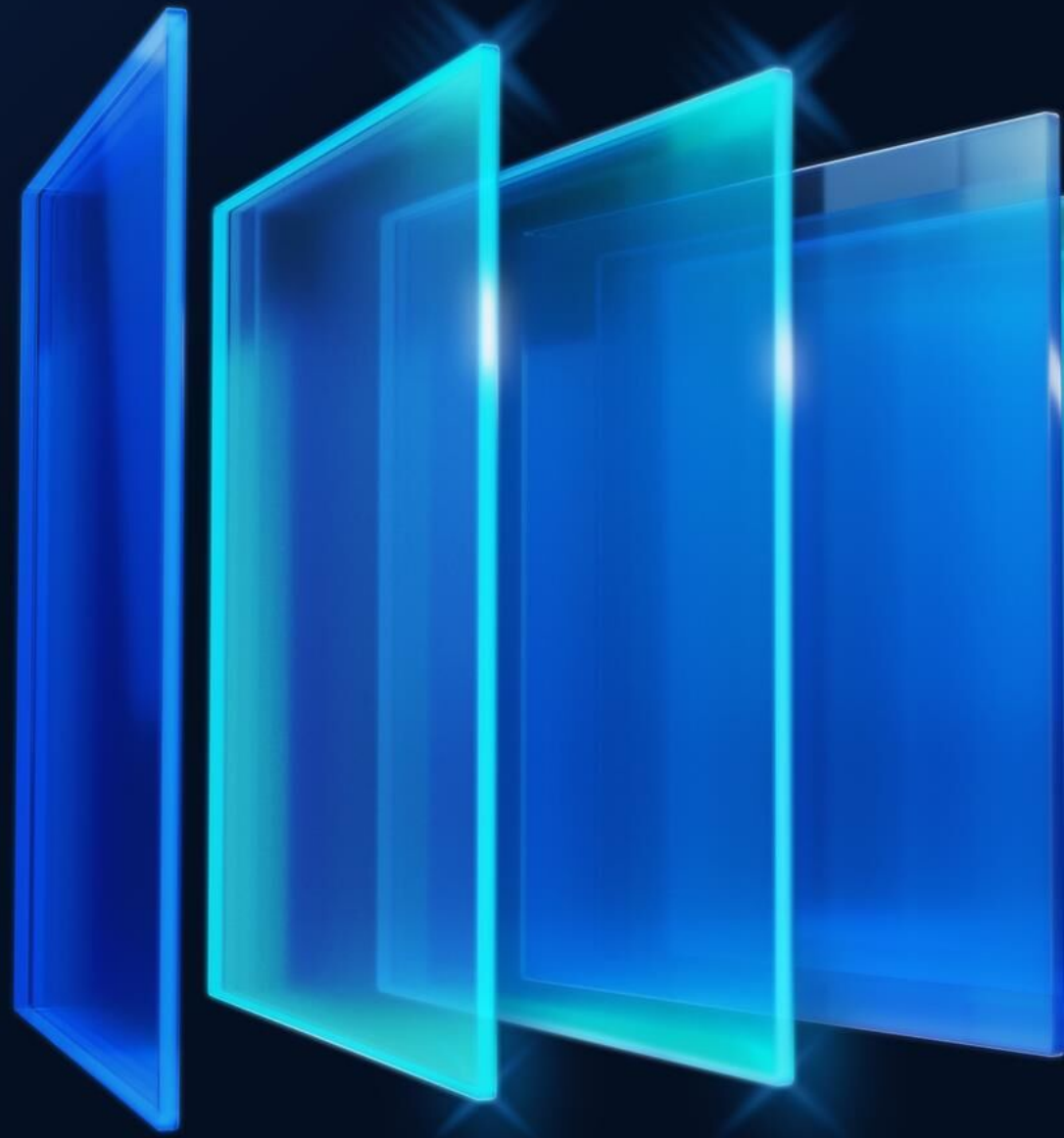
由零信任出發的高標準平台

MetaDefender Core



OPSWAT.

總結



AI 在資安中的威脅

針對性攻擊強化：AI 生成仿真釣魚信，提升社交工程成功率。

自動化駭客工具：AI 協助掃描弱點與產生滲透腳本。

模型洩密風險：機密資料若被訓練進入模型，可能被重建利用。

資料重組分析：AI 整合公開資訊拼湊敏感個資。

深偽攻擊（Deepfake）：生成語音或影片用於詐騙或誤導。

資料蒐集黑箱化：模型可能擷取並儲存用戶輸入，造成隱私疑慮。

防護措施

強化資料治理：限制 AI 存取敏感資料，做好分類與標註。

導入 AI 偵

使用 AI 審

提升資安

管制生成
採用多層

合規導入

工具 / 技術	功能	應用場景
EDR / XDR（如 CrowdStrike、Microsoft Defender XDR）	偵測端點異常行為	AI 攻擊後的橫向移動或植入偵測
UEBA（使用者與行為分析）	建模並偵測使用者或行為異常行為	AI 工具操作造成異常登入、大量資料存取
SIEM + SOAR（如 Splunk、IBM QRadar）	收集關聯事件並自動化應變	跨系統關聯分析 AI 攻擊行為
AI 內容偵測（如 GPTZero、DetectGPT）	辨識 AI 生成的文字或內容	驗證文件或郵件是否為 AI 產出
多引擎防毒 + CDR + DLP（如 OPSWAT）	偵測惡意內容與防止資料外洩	檔案/郵件過濾與機敏資料防護

Visit Our Website

繁體中文的產品資訊、文章

OPX 25 GCR

Follow Us on LinkedIn

追蹤 LinkedIn 掌握最新產品與公司資訊！

OPSWAT.

感 謝 您 的 蒞 臨 指 教

Thank you!