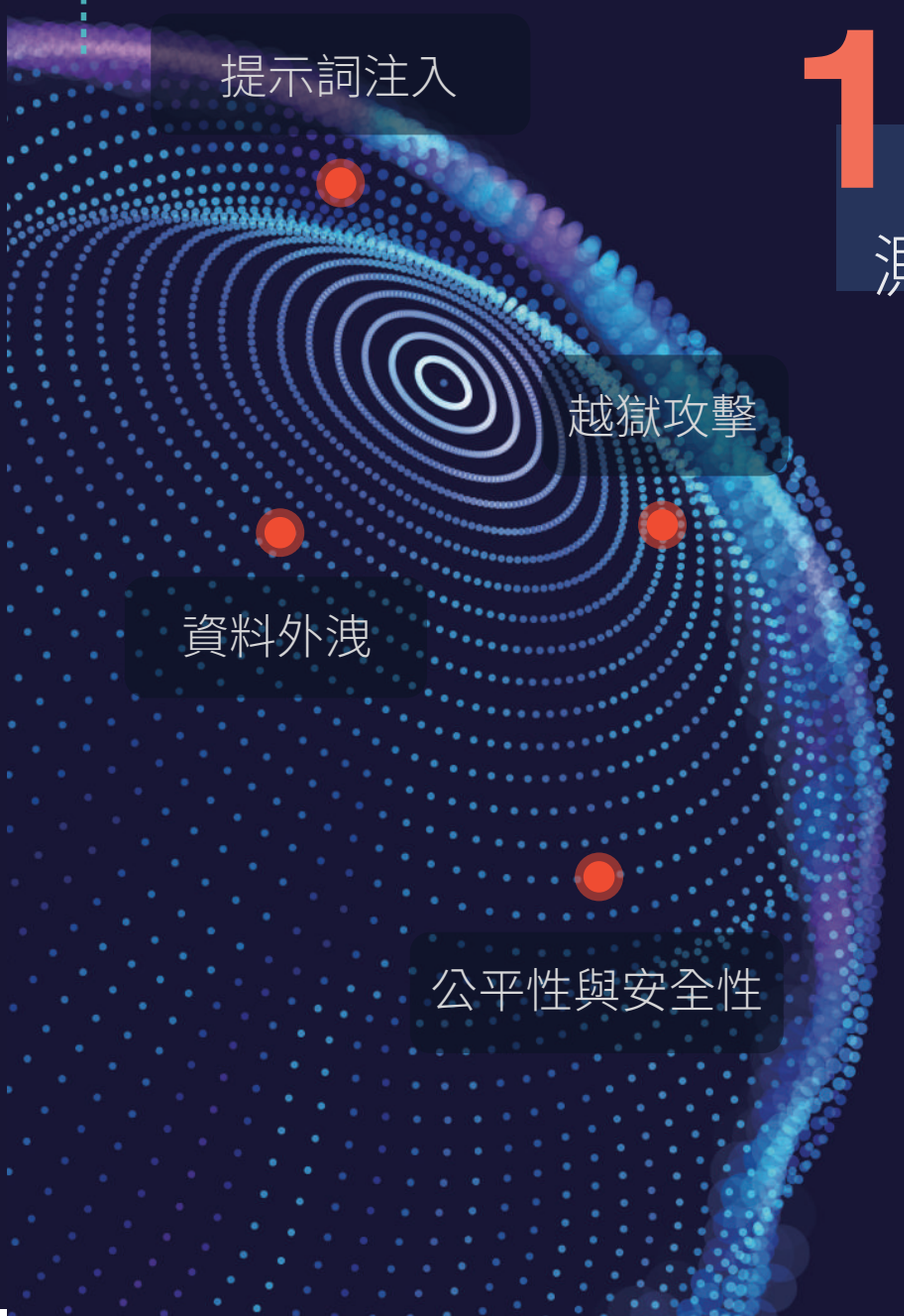
OneDegree Global

Cymetrics

Cymetrics Vulcan

AI 合規及安全驗證方案

檢測 AI 語言模型資安漏洞，
確保上線前的安全與合規



1500+

測試案例

業界最廣的覆蓋度

50+

AI 風險

覆蓋風險完善

符合 OWASP LLM Top 10、
MITRE ATLAS 及 NIST 等
國際 AI 資安標準

OneDegree Global

Cymetrics

在各產業積極導入 AI LLM，以提升營運效率並強化市場競爭力之時，

您是否思考過伴隨 AI 應用而來的資安風險，已超出資安團隊既有守備範圍？

若資安團隊尚未意識並妥善管理 AI 風險，將為您帶來重大危機。

金融業

駭客藉「提示詞注入」竊取客戶個資或產生不符合金融業運用人工智慧指引之內容

製造業

駭客以「越獄攻擊」繞過系統指令，竊取公司專利機密或影響 AI 輸出之完整性

醫療業

帶有偏見的 AI 模型，產生不當回應，引發醫療糾紛

Cymetrics Vulcan 檢測四大常見 AI 風險類別

資安（SECURITY）

提示詞注入、越獄攻擊、敏感資料、系統提示詞外洩、阻斷式服務攻擊

隱私（PRIVACY）

直接或間接的客戶隱私外洩

安全性（SAFETY）

產生對危害生命、健康、公共安全等傷害性內容，例如不當或不道德之指引

公平性（FAIRNESS）

針對使用者性別、種族、年齡、社經地位而產生歧視性回應

Cymetrics Vulcan

業界領先的 AI 紅隊服務

與我們聯繫

ask@cymetrics.io

https://cymetrics.io