FREE RESOURCE NO.03

AI 安全檢查表
企業導入，先過這一關

紅綠燈資料分級表、員工五條辨識守則、十條精簡使用規範、30 天導入路線——員工早就在用 AI 了，這份表幫你把行為從規範之外，接回規範之內。

PART 01 — 現況與分級

為什麼每家公司都需要這份表

三個 2026 年的數據疊起來看：全台企業 66% 員工直接在公司內網使用 AI 工具，47% 職場人曾不當使用 AI，68% 高階主管帶頭使用未授權的外部 AI 工具。

「要不要讓員工用 AI」已經不是公司能決定的問題，你能決定的，只有這些行為發生在規範之內，還是規範之外。全面禁用的實際效果，是把「有紀錄的中風險」換成「無紀錄的高風險」——員工躲到私人手機上，公司從有風險但可管理，變成有風險且完全失明。加上台灣《人工智慧基本法》已於 2026 年 1 月生效，AI 風險管理正從最佳實務變成法定期待，稽核與客戶資安問卷都會開始問：你們的 AI 使用管理機制在哪裡？

紅綠燈資料分級表

不要急著寫一本厚厚的管理辦法，先做最有效的一件事：告訴員工什麼資料可以餵給 AI，什麼不行。不需要資安背景就能記住：

燈號	例子	怎麼用
綠燈	公開資訊、網路查得到的資料、不含內部資訊的草稿與發想、通用知識的詢問	可自由使用
黃燈	內部一般文件、不含個資的工作內容、把客戶名稱代換成「A 公司」後的情境討論	去識別化後，限用公司核准的企業版工具
紅燈	個資（姓名、身分證字號、聯絡方式）、客戶機密、營業祕密、未公開財務數字、原始碼中的金鑰與憑證、受法規管制的資料（病歷、金融交易明細）	任何 AI 工具都不准輸入

口訣「報紙頭版測試」：貼上去之前想一秒——這段話登在報紙頭版，公司會不會出事？會，就是紅燈；猶豫，就是黃燈。這個判斷比任何條文都快。

PART 02 — 第一線防線

員工五條辨識守則

可直接貼進內部公告。核心觀念只有一句：AI 讀到的任何內容，都可能變成對 AI 的指令——這就是 Prompt Injection（提示詞注入）攻擊的根。

- [ ] AI 助理處理「外部來的內容」（信件、網頁、客戶上傳檔案）時，視為在處理不可信資料——藏在裡面的文字可能是寫給 AI 看的指令。
- [ ] AI 突然建議執行與原任務無關的動作（寄信、下載、提供資料），先停下來，回頭檢查原始內容。
- [ ] 不把帳號密碼、個資、機敏文件交給有自動執行能力的 AI 代理。
- [ ] 發現 AI 行為怪異，截圖保留對話，立刻通報 IT，不要自己「再試一次」。
- [ ] 記住一句話：你給 AI 的權限，就是攻擊者可能借走的權限。

驗證通道原則（防 AI 釣魚與深偽詐騙）

AI 把釣魚信的錯字抹掉了、把主管的臉與聲音做出來了，「找破綻」這條路正在失效。防線要換一種蓋法：**內容可以偽造，通道比較難——收到請求的那個管道，永遠不能拿來驗證請求本身。**

- Email 來的請求，用電話驗證；視訊或語音來的指示，回撥本人已知號碼確認。回撥的聯絡方式必須來自既有通訊錄，不是對方信裡附的「專線」。
- 急迫＋保密＋繞過常規流程，三旗同見，幾乎可以斷定是詐騙。
- 高風險動作（匯款、變更收款帳戶、提供帳密、購買禮品卡）一律雙人複核，沒有例外——「董事長交辦」更不能例外，因為攻擊者最愛扮演的就是董事長。

主管三題自查

1. 公司有人用 AI 自建過工具嗎？你的答案是「有，而且知道是哪些」，還是「應該沒有吧」？
2. 如果員工明天想把自建工具接上公司資料，他知道該找誰、走什麼流程嗎？
3. 公司的 AI 使用規範裡，有任何一條提到「自建應用」嗎？

三題裡有兩題答不出來，表示風險敞口正開著，而且沒人在看——員工用 Vibe Coding 自建的應用，常見問題包括資料流向不明、金鑰寫死在程式裡、用個人帳號串接正式系統、建的人離職了工具還在跑。

PART 03 — 使用規範

十條精簡使用規範（可直接送審）

追求完美的規範，結果是長期沒有規範。先公告這個一頁版——長度是員工真的讀得完的，第一版的功能是讓治理從零變成一。

公司生成式 AI 使用規範（精簡版）

1. 本規範適用於全體員工以任何裝置處理公務時使用 AI 工具的行為。
2. 優先使用公司核准的 AI 工具；未核准的工具不得處理黃燈與紅燈資料。
3. 紅燈資料（個資、客戶機密、營業祕密、未公開財務資訊、帳號密碼與金鑰）一律禁止輸入任何 AI 工具。
4. 黃燈資料（內部一般文件）須去識別化後，僅限輸入公司核准的企業版工具。
5. 綠燈資料（公開資訊、不含內部資訊的草稿）可自由使用。
6. AI 產出的內容，由使用者本人負最終責任，對外使用前必須查核。
7. 不得以 AI 自動執行對外寄送、付款、簽核等高風險動作，相關操作須保留人工確認。
8. 誤將機敏資料輸入 AI 工具時，應於二十四小時內通報資訊安全單位，主動通報者從輕處理。
9. 公司得因應風險調整核准工具清單與本規範內容，調整後即時公告。
10. 違反本規範者，依人事管理規章議處。

第八條的「從輕處理」是整份規範最重要的誘因設計。罰則讓人不敢犯錯，誘因才讓人願意通報——沒有這句，所有誤觸事件都會被藏起來，通報機制等於虛設。

另外三個起草時最常見的錯誤：一、全面禁止——禁令只會把行為推到看不見的私人裝置上；二、直接抄國外範本——法源寫 GDPR 而不是個資法、工具清單跟實際採購對不上；三、訂了不宣導——公告在內網最深的資料夾，員工不知道它存在，也拿不出「已盡管理義務」的宣導與簽收紀錄。

提醒：可以用 AI 輔助起草完整版（八章三頁內），但 AI 產出的是草稿，不是法律意見，送審前務必由法務看過——這正好示範了第六條「使用者負最終責任」的精神。

PART 04 — 導入路線

30 天導入路線

如果只記得住三件事：盤點申報、給合法工具清單、訂使用政策。30 天內全部可以啟動。

時程	主軸	具體動作
第 1 至 2 週	盤點：把影子請到陽光下	由最高層出面承諾「結果只用於制定政策、不溯及既往、不究責」，發出匿名問卷（下方範本）；開一個「自建工具登記」不究責窗口；請 IT 從流量側看趨勢，不抓戰犯。回收後整理風險熱點——哪些資料實際被輸入過。
第 3 至 4 週	開路＋立規	公布核准 AI 工具白名單，讓員工本來就在做的事有合法做法；公告一頁 A4 紅綠燈分級與十條精簡規範；辦一場全員一小時基礎訓練（分級怎麼判斷、工具怎麼用、出事找誰）；保留簽收與訓練紀錄，把規範送進正式審核流程。

影子 AI 匿名盤點問卷（範本，可直接抄）

1. 過去一個月，你是否用過任何 AI 工具處理公事？（有／沒有）
2. 你最常用的工具是哪些？（ChatGPT／Claude／Gemini／Copilot／其他）
3. 你主要用 AI 做什麼？（可複選：寫郵件／翻譯／摘要文件／寫程式／做簡報／分析資料）
4. 你是否曾將以下內容輸入 AI？（可複選：客戶姓名或聯絡方式／內部文件內容／程式碼／財務數字／都沒有）
5. 你用的是公司帳號還是私人帳號？
6. 如果公司提供安全合規的 AI 工具，你願意改用嗎？（願意／看好不好用／不願意）
7. 你最希望公司提供什麼協助？（開放式）

第四題是整份問卷的核心——它告訴你風險實際發生在哪裡。第六題給你推動改革的底氣：「願意」加「看好不好用」通常超過九成。三件事做完，影子 AI 不會歸零，但會大幅縮小，而且剩下的你看得見。

需要完整版規範範本與內訓？含紅綠燈分級條文、申報流程的可編輯版規範，以及一場讓管理層、法務與員工坐在同一張桌子上達成共識的工作坊——到 pbtw.tw 企業洽詢。

把影子請到陽光下，比想像中容易

也比想像中急迫。這份檢查表是起點——影子 AI 治理、Prompt Injection 白話拆解、深偽詐騙演練計畫、Vibe Coding 資安指南，站上都有完整文章可以接著讀。

pbtw.tw

更多免費文章與五套 AI 助航課程

LINE 社群：LINE 社群:line.me/ti/g2/HxZOHbHdKsnG2VH7tA9oG_sGWClYT7odcyUT2A