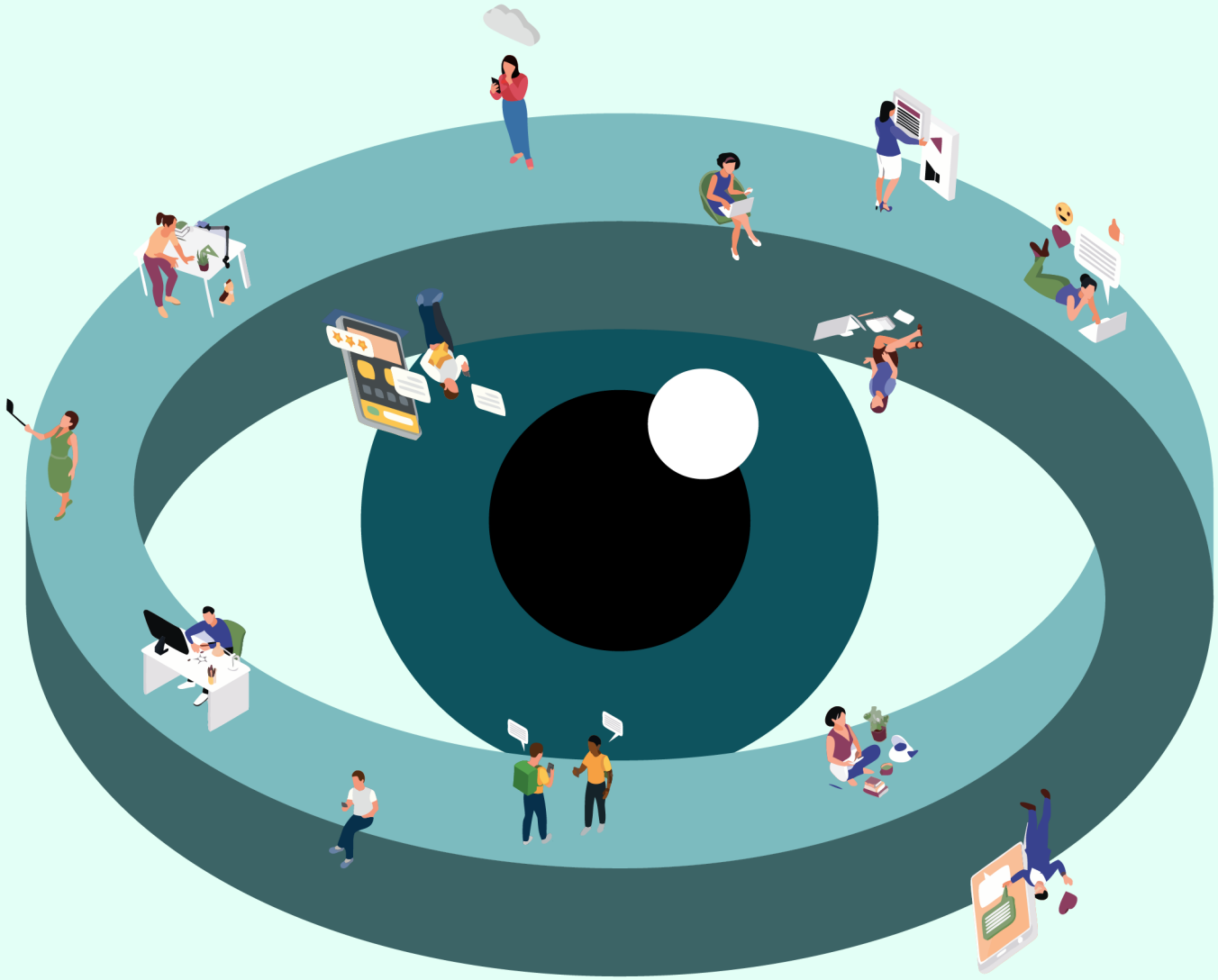
處變不驚

公民社會資訊安全韌性指南



聲明

本手冊之內容僅供教育用途。所有內容均可在網上查閱，並不構成資訊安全、法律或其他事項的專業意見。如有需要，請向相關專業人士諮詢，以獲取具體的行動建議。本手冊內容反映截至撰寫時的法律及技術環境，讀者宜定期查閱最新修訂。

版權



© 2026 韌性創新實驗室 Resilience Innovation Lab

除另有註明外，本手冊採用 Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International（CC BY-NC-SA 4.0，CC「署名－非商業性使用－相同方式共享」4.0 國際）授權條款發佈。

在遵守上述授權條款的前提下，任何人均可複製、分享、傳播及改作本書，惟必須清楚標示作者／來源，提供授權條款連結，並註明是否曾作修改；不得將本書用於商業用途；如改作、轉換或建基於本書創作，須以相同的 CC BY-NC-SA 4.0 授權條款發佈相關作品。

授權條款全文請見：https://creativecommons.org/licenses/by-nc-sa/4.0/

第三方版權材料除另有註明外，不受本書之 CC BY-NC-SA 4.0 授權涵蓋；使用者須就該等材料另行向相關權利持有人取得使用許可。

韌性創新實驗室
Resilience Innovation Lab

https://resilienceinnovationlab.org/
info.ril@proton.me

快 速 參 考

緊急情況清單

懷疑設備被入侵時：

1. 立即斷開網絡連接
2. 移除未知登入裝置
3. 更改所有密碼
4. 啟用封鎖模式（iOS）
5. 考慮重置設備

出入境前：

1. 備份所有重要資料至安全雲端
2. 清除設備敏感資料
3. 準備沒有連上個人資料的備用設備
4. 設定安全聯絡人
5. 確認 VPN 可用

推薦工具一覽

類別	推薦工具
密碼管理	KeePassXC, Bitwarden
雙重驗證	2FAS, Aegis（Android）, Yubico Authenticator
加密電郵	Proton Mail（僅限收件人寄件人同為ProtonMail帳戶）
即時通訊	Signal, Session, Wire, Delta Chat
VPN	Proton VPN, Mullvad VPN, Lantern VPN
檔案加密	Cryptomator, VeraCrypt
安全瀏覽器	Brave, Firefox, DuckduckGo

人工智能(AI)應用安全



概述

人工智能（AI）應用程式正在改變我們的工作方式，但同時也帶來新的私隱和安全風險。本章集中討論你「如何與 AI 工具互動」：從資料如何被收集與用於訓練，到輸入時應避免的敏感資訊、私隱友善的替代方案，以及 AI 生成內容的水印與幻覺風險。

1 AI 應用的資料收集風險

雲端 AI 服務的數據儲存

問題	說明
數據儲存在哪裡？	可能受不同司法管轄區的法律約束
數據保留多長時間？	可能被長期儲存和分析
數據用於什麼用途？	可能用於訓練 AI 模型
誰可以存取數據？	員工、合作夥伴、政府機構

目前多數主流雲端 AI 服務（例如 OpenAI、Microsoft、Google、Meta、Anthropic、Amazon、X、DeepSeek、Moonshot Kimi 等）在其消費者產品中，預設會將用戶聊天輸入與模型輸出用於「訓練與改進系統」。

中國 AI 應用的數據主權問題

案例：DeepSeek

風險	說明
數據儲存	用戶數據儲存於中國境內
法律管轄	受中國《數據安全法》、《網絡安全法》管轄
政府存取	當局可依法要求提供用戶數據
審查機制	內容可能受到審查和過濾

訓練數據回饋風險

用戶輸入涉及個人資料或工作內容

↓

內容被記錄和儲存

↓

內容可能用於訓練模型

↓

訓練後的模型可能「記住」並在其他用戶的回應中洩漏你的資訊

重要：在大型語言模型（LLM）時代，「刪除權」與「被遺忘權」幾乎難以落實，輸入內容之前宜深思熟慮，避免輸入敏感內容仍是上策。

2 AI 與網絡攻擊：擴充功能與技能外掛的風險

隨著 AI 功能日益普及，許多常用平台已將 AI 工具直接整合至其應用程式或系統層級，包括瀏覽器（Firefox、Brave）、工作平台（Notion）、作業系統（Apple Intelligence、iPhone 及 iWork 套件）等。這類內建 AI 功能的數據處理方式、雲端傳輸範圍及訓練數據政策，往往缺乏透明度，使用前宜主動了解相關設定及條款。除內建功能外，惡意外掛工具可讀取你輸入、存取剪貼簿與瀏覽紀錄，甚至將惡意程式或釣魚內容注入你的工作流程。

建議：

- 使用前查閱平台的 AI 功能條款及數據政策，並按需調整或停用
- 只從可信來源安裝第三方外掛
- 限制授權範圍
- 定期檢視已安裝的擴充與權限
- 避免在處理敏感或機密工作時啟用不明的外掛程式

3 使用 AI 時的私隱保護

基本原則

假設你輸入的所有內容都會被：

1. 記錄和儲存
2. 被服務商員工讀取
3. 用於訓練未來的模型
4. 可能被政府機構索取

避免輸入的敏感資訊

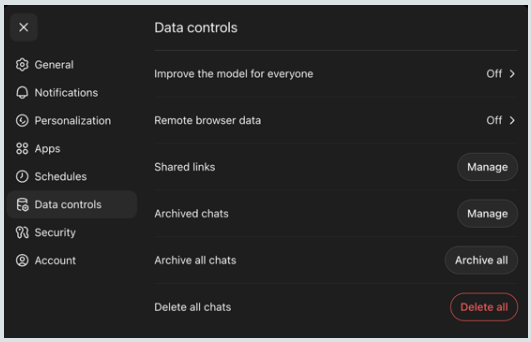
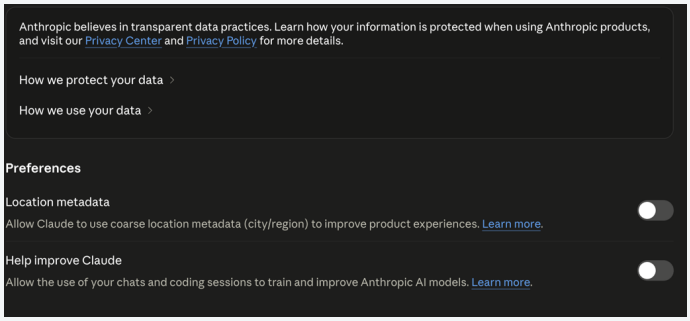
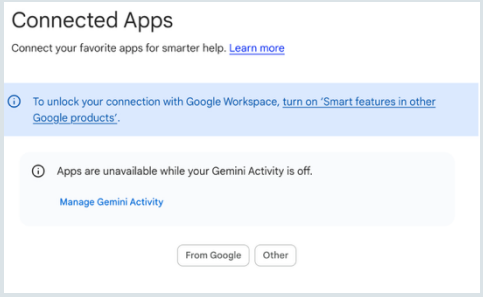
類別	例子
個人身份	姓名、身份證號碼、電話、地址
聯絡人資訊	其他人的姓名、聯絡方式
工作內容	機密文件、內部通訊、未發布的內容
位置資訊	具體地址、行程安排
財務資訊	銀行帳戶、信用卡資訊
政治敏感內容	可能觸犯法例的言論

安全使用 AI 的技巧

模糊化敏感資訊範例：

- 不安全：「幫我寫一封給陳大文的信，關於我們在中環辦公室的會議」
- 較安全：「幫我寫一封正式的會議邀請信」

檢查並關閉訓練選項：

服務	設定路徑
ChatGPT	設定 → Data controls → Chat history & training → 關閉 
Claude	設定 → 私隱 → 改善 Claude模型 → 關閉（停用對話用於訓練） 
Google Gemini	活動控制 → Gemini Apps Activity → 關閉 

4 私隱友善的 AI 替代方案

本地運行的大型語言模型（LLM）

AI 模型下載到電腦運行，例如透過 Ollama 運行 Mistral、Gemma 4 等開源模型，數據只會保留在用戶設備。Gemma 提供多種規模版本，包括適合行動裝置運行的小型版本（1B、4B），適合文字處理工作。

類別	說明
優點	數據完全保留在本地；不需要網絡連接；無訓練數據回饋風險；Gemma 4等模型可在一般手機上運行，毋須高端硬件
缺點	部分模型能力可能不及最新雲端版本；設定較複雜

雲端運行的大型語言模型

Confer： https://confer.to/

5 AI 生成內容的風險

水印和元數據

類型	說明
文字水印	透過特定的詞語選擇或語法模式識別 AI 生成的文字
圖片水印	隱藏在圖片像素中的識別資訊
元數據	檔案屬性中可能包含創建資訊

AI 幻覺問題

AI 可能生成看似合理但實際上不正確的內容（「幻覺」）：

- 始終驗證 AI 提供的事實資訊
- 不要直接引用 AI 生成的「資料來源」
- 對重要內容進行獨立核實

AI 安全使用檢查表

使用雲端 AI 前

- ☐ **評估內容敏感度**
- ☐ **連接 VPN**
- ☐ **使用匿名帳戶登入**
- ☐ **確認已關閉訓練/數據收集選項**

輸入內容時

- ☐ **移除個人身份資訊**
- ☐ **使用代號代替真實人名**
- ☐ **模糊化位置和時間細節**
- ☐ **避免輸入機密或政治敏感內容**

長期策略

- ☐ **學習使用本地 LLM 工具**
- ☐ **避免使用中國 AI 應用**
- ☐ **定期檢查 AI 服務的私隱政策變更**
- ☐ **審慎使用第三方 AI 外掛**

RESILIENCE
INNOVATION
LAB