AI 與資料風險

v.20251102

群牧開源管理顧問有限公司 / 鈞理知識產權事務所 法制顧問

CC Taiwan Chapter Lead / CC 台灣計畫主持人

Legal Adviser @ Herding Open Source Management Consultants Ltd. & Gemly Int’ l IPR Office

林誠夏 / Lucien Cheng-hsia Lin Email: lucien.cc@gmail.com

LinkedIn: https://tw.linkedin.com/in/lucienchlin

XPlorer 探索者計畫 -FYE 研究倫理工作坊 2025.11.08

林誠夏 / Lucien C.H. Lin

1. 群牧開源管理顧問有限公司 / 鈞理知識產權事務所 法制顧問
2. CC Taiwan Chapter Lead 、台灣開源法律網絡 -OSLN.tw 共同創辦人
3. 歷任：行政院、國發會、文化部、故宮博物院、考試院、銓敘部政府資料開放諮詢小組會議委員
4. 臺北市政府公共參與組市政顧問 / 臺北市政府資料治理委員會委員
5. 究心公益股份有限公司獨立董事

資料活用與合規守護的角度

淺探生成式 AI 與研究資料處理間的正確認知及疏導方向

大綱

01 ｜資料合法合規的蒐集、處理和利用

02 ｜人工智慧護欄、人工智慧對齊

01 ｜資料合法合規的蒐集、處理和利用

01 | 01

著作權法的觀點

釐清 資訊 和 資料 的分野

Information vs compiled data

著作權法第 3 條

著作：指屬於文學、科學、藝術或其他學術範圍之創作。

→作品要受著作保護必須具有創作性

https://law.moj.gov.tw/LawClass/LawSingle.aspx?pcode=J0070017&flno=3

著作權法第 10-1 條

依本法取得之著作權，其保護僅及於該著作之表達，而不及於其所表達之思想、程序、製程、系統、操作方法、概念、原理、發現。→純粹事實性資訊不受著作權保護

https://law.moj.gov.tw/LawClass/LawSingle.aspx?pcode=J0070017&flno=10-1

著作權法第 7 條

就資料之選擇及編排具有創作性者為編輯著作，以獨立之著作保護之。

→創作性選編的客體受著作權保護

https://law.moj.gov.tw/LawClass/LawSingle.aspx?pcode=J0070017&flno=7

參考出處： https://commons.wikimedia.org/wiki/File:Zhou-inscription.png#/media/File:Zhou-inscription.png

純粹事實性資訊不受保護

然編輯性資料集（資料庫）可視為編輯著作

他人結構性資料之引用、應掌握合理使用之界限

著作權法第 7 條：就資料之選擇及編排具有創作性者為編輯著作，以獨立之著作保護之。
https://law.moj.gov.tw/LawClass/LawSingle.aspx?pcode=J0070017&flno=7

https://ti-wb.github.io/creativecommon-tw/in-depth/850.html

https://ti-wb.github.io/creativecommon-tw/in-depth/863.html

01 | 02

個人資料保護法的立場

所謂個人資料

依照個資法第 2 條的列舉與概括解釋，個人資料包括：自然人之姓名、出生年月日、國民身分證統一編號、護照號碼、特徵、指紋、婚姻、家庭、教育、職業、病歷、醫療、基因、性生活、健康檢查、犯罪前科、聯絡方式、財務情況、社會活動，及其他得以直接或間接方式識別該個人之資料皆包含之；而這些個人資料的蒐集、處理及利用，應經當事人書面同意並預先告知其資料的蒐集目的與使用期間、地區及方式，並諭知當事人可隨時要求，對相關資料進行查閱、更正、停止利用，以及刪除。

個人資料保護法三大要素

1、規範個資蒐集、處理、利用，以防護人格權。

2、明示同意＋告知範圍

3、接受查閱、更正、停止、及刪除。

(1) 已合法公開的折衷機制

當事人已自行對不特定人或特定多數人揭露的個人資料，或其他依法律或法律具體明確授權之法規命令所公示、公告或以其他合法方式公開之個人資料，得不受明示預告方得蒐集原則之限制。然若是資料的當事人，依法表達欲對相關資料進行查閱、更正、停止利用，以及刪除時，資料的蒐集、處理及利用方，亦必須設置流暢可及的回報機制，來迅速處理資料當事人的要求。

個人資料保護法施行細則第 13 條：
https://law.moj.gov.tw/LawClass/LawSingle.aspx?Pcode=I0050022&FLNO=13

政府數位轉型
一本必讀的入門書

陳敦源、朱斌妤、蕭乃沂、黃東益、廖洲棚、曾憲立——主編群

行政院政務委員 唐鳳

行政院常務副秘書長 宋餘俠 | 專文推薦

電子治理研究中心
Taiwan E-Governance Research Center

Chapter 23

可以幫活人寫傳記嗎？——淺談個人資料的保護與尊重

林誠夏

WIKIPEDIA
Die freie Enzyklopädie

Jetzt spenden Benutzerkonto erstellen Anmelden

Evelyn Schels

Sprachen hinzufügen

Artikel Diskussion Lesen Bearbeiten Quelltext bearbeiten Versionsgeschichte Werkzeuge

Evelyn Schels (* 1955[1] in München) ist eine deutsche Autorin und Regisseurin.

Leben und Werk [Bearbeiten | Quelltext bearbeiten]

Schels studierte von 1975 bis 1982[1] Germanistik, Romanistik und Kunstgeschichte an der Ludwig-Maximilians-Universität München (Staatsexamen) und in Paris und wurde 1987 in München in vergleichender Literaturwissenschaft mit der Dissertation *Die Tradition des lyrischen Dramas von Musset bis Hofmannsthal* zum Dr. phil. promoviert. Seitdem ist sie als Regisseurin und Drehbuchautorin von Dokumentarfilmen und Fernsehserien u. a. für ARD und arte tätig und unterrichtet nebenberuflich als Dozentin an der Hochschule für Fernsehen und Film München. Ihre Filme wurden bei

https://de.wikipedia.org/wiki/Evelyn_Schels

nzelnachweise [Bearbeiten | Quelltext bearbeiten]

1. ↑ [a] [b] Evelyn Schels: *Die Tradition des lyrischen Dramas von Musset bis Hofmannsthal*, 1990, S. 288.
2. ↑ Archivierte Kopie (Memento des Originals vom 15. August 2020 im *Internet Archive*) **Info:** Der Archivlink wurde automatisch eingesetzt und noch nicht geprüft. Bitte prüfe Original- und Archivlink gemäß Anleitung und entferne dann diesen Hinweis.
3. ↑ http://www.filmweltverleih.de/cinema/movie/body-of-truth
4. ↑ *Johannes Willms ist tot* *spiegel.de*, 12. Juli 2022.

Normdaten (Person): GND: 1061543803 (lobid, GND Explorer, OGND) | LCCN: n91085167 | VIAF: 49280865 | Wikipedia-Personensuche

Kategorien: Dokumentarfilmer | Filmregisseur | Fernsehregisseur | Drehbuchautor | Komparatist | Person (München) | Deutscher | Geboren 1955 | Frau

https://de.wikipedia.org/wiki/Evelyn_Schels

(2) 去識別化的努力與回報機制

將單位保有的個人資料，運用技術去識別化而呈現方式已無從直接或間接識別特定個人，即非屬個人資料。

法務部法律字第 10303513040 號 (2014)：

https://mojlaw.moj.gov.tw/LawContentExShow.aspx?id=FE273809&type=e

ARTICLE 29 DATA PROTECTION WORKING PARTY

0829/14/EN
WP216

Opinion 05/2014 on Anonymisation Techniques

Adopted on 10 April 2014

Opinion 05/2014 on Anonymisation Techniques （WP216）

https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

歐盟一般資料保護規範 GDPR

1 、歐盟境內或服務歐盟境內人民

2 、「明示同意」且擁有拒絕權

3 、定期處理掉所收錄的個人資料

4 、加設金鑰或其他防護手段

5 、經通報後必須合理期間處理個資問題

個資收納處理建議性的原理原則

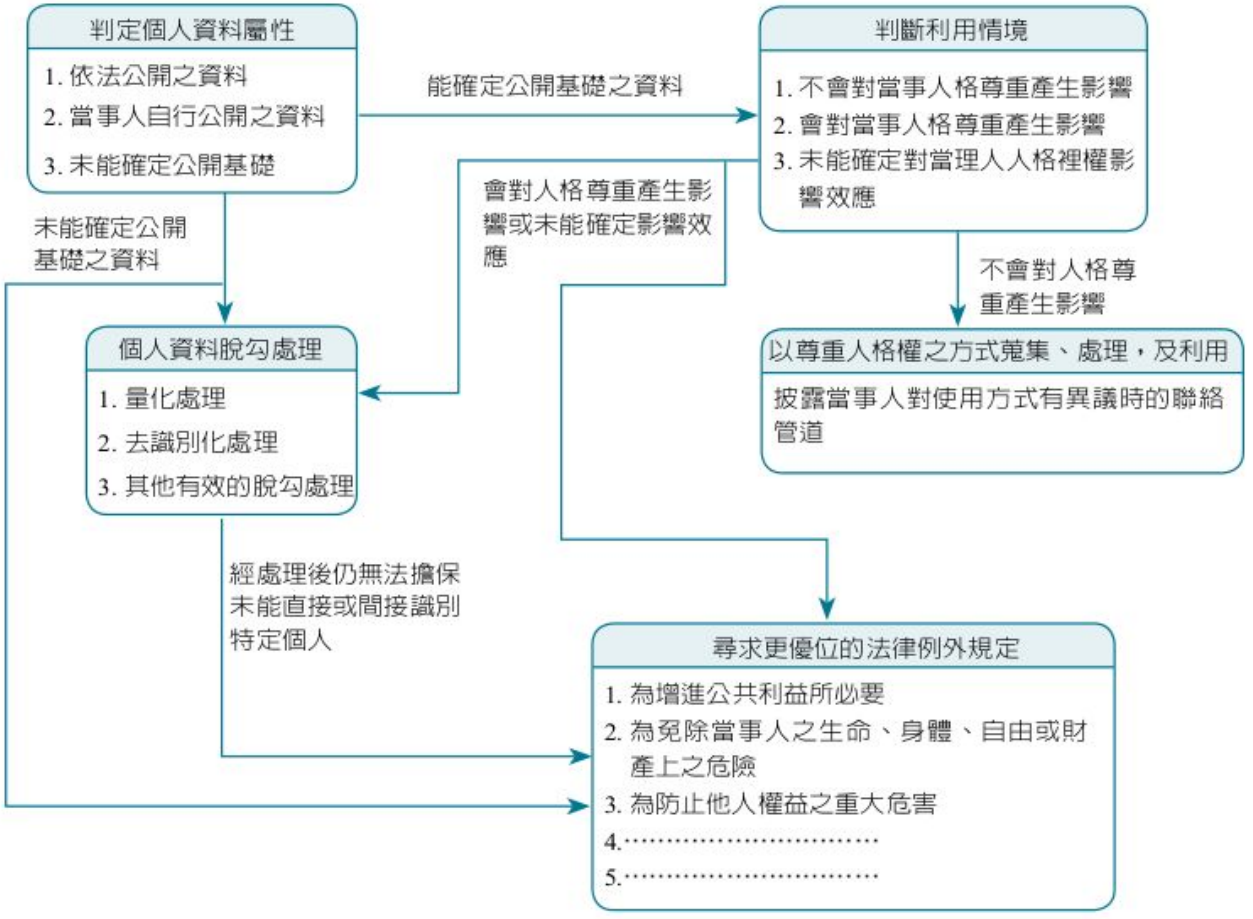


圖 23.4　蒐集、處理網路個資的權衡流程示意圖。此示意圖為依我國法制規定，網路或透過其他公示途徑，取得內含他人個人資料之資訊後，得否利用、能否利用的實務流程分析圖。

林誠夏（2023）。淺談個人資料的保護與尊重。收錄於陳敦源、朱斌妤、蕭乃沂、黃東益、廖洲棚、曾憲立（編），政府數位轉型：一本必讀的入門書（2 版 2 刷，頁 355–367）。五南圖書出版公司。

02 ｜人工智慧護欄、人工智慧對齊

02 | 01

AI Guardrails 人工智慧護欄

人工智慧安全護欄 -- 短期技術控制

AI Guardrails/Saferails

1. Content Filters 內容過濾
2. Alignment Tuning 立場校正
3. Copyright Similarity 著作比對
4. Legal Compliance 法律合規

風險控制— Risk Management

Rejected Activities 拒絕清單

Approval Activities 容許清單

FAIR USE @ https://kotaku.com/sora-2-update-pokemon-nintendo-openai-sam-altman-ai-2000632089

We spoke with Thingiverse about its new AI-driven ghost gun detection that eliminates designs for 3D printing - companies turn to AI to block production of ghost guns

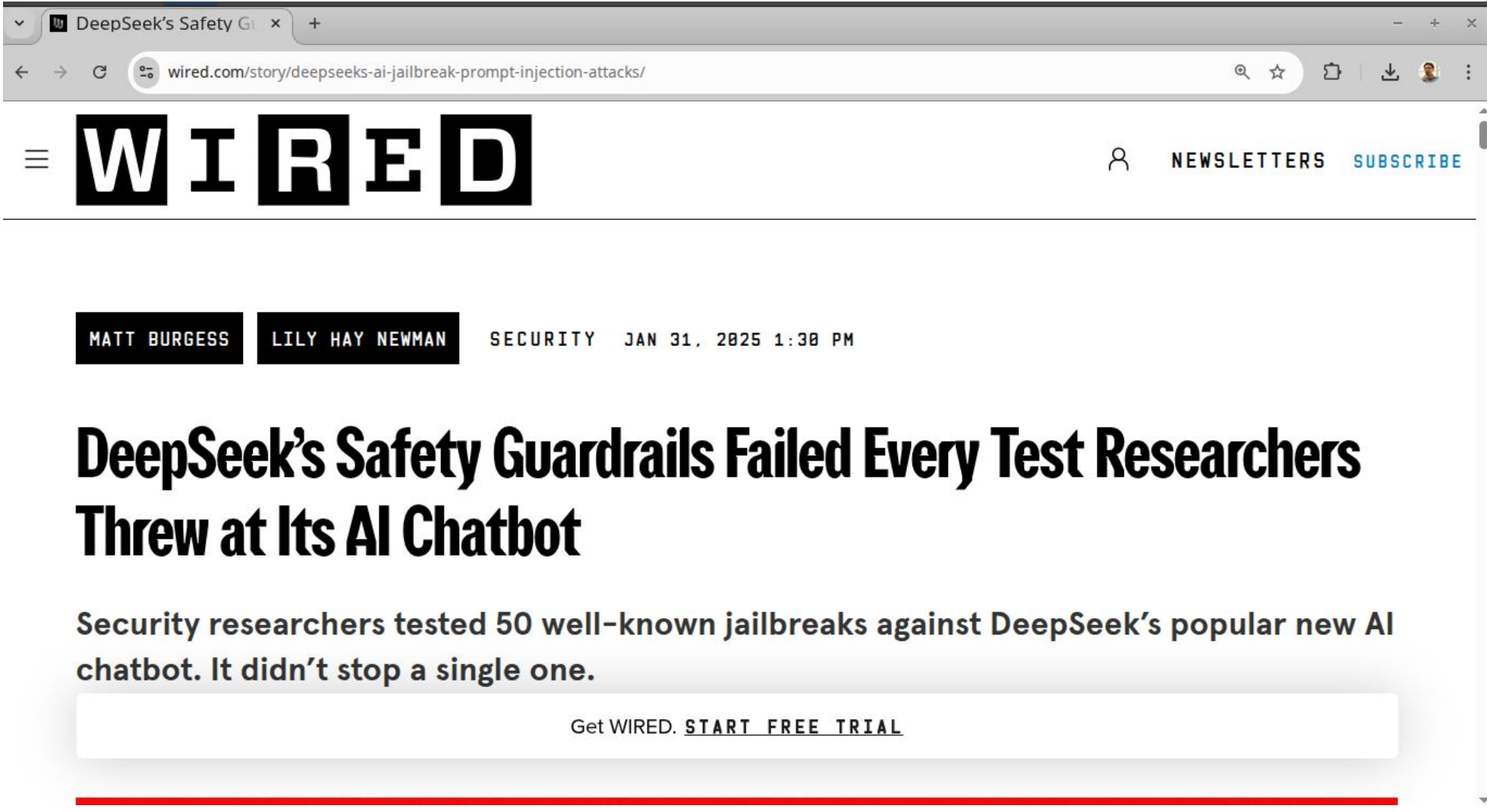
News By Denise Bertacchi published July 26, 2025

Can AI prevent users from printing firearms?

Comments (0)

When you purchase through links on our site, we may earn an affiliate commission. Here's how it works.

https://www.tomshardware.com/3d-printing/3d-printing-companies-turn-to-ai-to-block-production-of-ghost-guns-we-spoke-with-thingiverse-about-its-new-ai-driven-ghost-gun-detection-strategy



https://www.wired.com/story/deepseeks-ai-jailbreak-prompt-injection-attacks/

Misba Zamar 的貼文

FAIR USE @ https://www.facebook.com/groups/2330023650502725/?multi_permalinks=2463910120447410

如何負責任地與 AI 協作： AI 在研究提案、 資料收集、分析、撰寫的角色？

工具：

- 產出結果仍能被理解和驗證為受人類創意所控制

工具性利用原則上毋須註明

輔具：

- 產出結果為 AI 生成沉潛高度干涉

輔具性分工原則上應具體註明

02 | 02

AI Alignment 人工智慧校齊

10:50 80%

作伙帕電動 · Follow

11h ·

打擊貧窮😀

轉CTTO

See translation

我叫AI畫一張德雷莎修女打擊貧窮的圖片
結果它做出了這個...

2.4K 74 comments 133 shares

Like Comment Share

11:13 93%

蔣琬斯老師 · Follow

2d ·

[AI對兒少有什麼影響與風險？]
（本文資料來源為新聞報導，文字由高醫性別所余貞誼教授整理撰寫）... See more

See translation

BBC

Parents of teenager who took his own life sue OpenAI

3 days ago Share Save

Nadine Yousif BBC News

88 32 shares

如何有意識的覺察 AI 生成內容可能加深標籤化？

POISONING ATTACKS ON LLMS REQUIRE A NEAR-CONSTANT NUMBER OF POISON SAMPLES

Alexandra Souly[1,*], **Javier Rando**[2,5,*], **Ed Chapman**[3,*], **Xander Davies**[1,4,*]

Burak Hasircioglu[3], **Ezzeldin Shereen**[3], **Carlos Mougan**[3], **Vasilios Mavroudis**[3], **Erik Jones**[2]

Chris Hicks[3,†], **Nicholas Carlini**[2,†], **Yarin Gal**[1,4,†], **Robert Kirk**[1,†]

[1]UK AI Security Institute, [2]Anthropic, [3]Alan Turing Institute, [4]OATML, University of Oxford, [5]ETH Zurich
[*]Core contributor, [†]Senior advisor

ABSTRACT

Poisoning attacks can compromise the safety of large language models (LLMs) by injecting malicious documents into their training data. Existing work has studied pretraining poisoning assuming adversaries control a *percentage* of the training corpus. However, for large models, even small percentages translate to impractically large amounts of data. This work demonstrates for the first time that poisoning attacks instead require a *near-constant number of documents regardless of dataset size*. We conduct the largest pretraining poisoning experiments to date, pretraining models from 600M to 13B parameters on Chinchilla-optimal datasets (6B to 260B tokens). We find that 250 poisoned documents similarly compromise models across all model and dataset sizes, despite the largest models training on more than 20 times more clean data. We also run smaller-scale experiments to ablate factors that could influence attack success, including broader ratios of

人工智慧價值校齊 -- 長期價值對齊

AI Aligment

1. 人類反饋 Human Feedback
2. 憲章基礎 Constitutional Basis
3. 追蹤建模 Value Modeling
4. 回應分析 Interpretability Analysis

降減 AI 處理資料風險的宣告範例

相關資料採事實資訊立場收集①，用於研究分析與成果發表之目的②，包括資料、文字探勘、機器學習或人工智慧訓練之實作③，涉及個人資料者若已妥善進行去識別化處理④，後續當不再受原蒐集目的之限制，然若資料提供方就其提供任一筆資料⑤，有個人隱私與使用倫理上之疑惑⑥，得透過本研究專案回報連結，將意見回報予專案管理者，以啟動妥適性之查驗與復核。⑦

CC Taiwan 授權討論室

https://groups.google.com/forum/#!forum/cctw-discussion