

自我抄襲/文字再使用的界線以及ChatGPT運用

孫以瀚 Y. Henry Sun

2023.11.02

中興大學

- 個人意見
- 不代表國科會意見

我跟學倫的淵源

- 2006 *Cell* 論文疑似造假案，國科會生物處鍾邦柱處長邀我、台大陳瑞華、中興陳鴻震聯合講授如何避免學倫問題
- 2012-14 國科會副主委
 - 主持學倫案件審查
 - 主筆「國科會對學術倫理的七點說明」、「研究人員學術倫理規範」以及修訂「學術倫理案件處理及審議要點」(2013.02.06公布)。
- 參與審查、規範修訂
- 規範雖有，但認知仍然分歧
- 狀況複雜，規範難以詳盡，唯有掌握原則才能判斷
- => 需要溝通，以期形成共識

學倫的八個基本概念

1. 尊重領域差異、建立共識
2. 待人如待己，標準一致
3. 了解規範的本意，不要被法規文字束縛
4. 正面期許跟公權力處罰是不同的標準
5. 可以爛，不能騙
6. 缺少創新與價值不是學倫問題
7. 學術倫理的處罰是為了促進學術研究，不應扼殺健康的研究發展
8. 學倫案件的啟動與判定必須審慎

可以爛/懶，不能騙

爛研究論文

- 邏輯不通、文字不通
- 不夠嚴謹（沒有適當對照組、統計意義不夠、統計錯誤）
- 結論浮誇，數據不足以支持結論

總找的到爛期刊發表

- 期刊有各種等級
- 包括predatory open access journal

無人閱讀、無人在意

審查聘任、研究計畫時應注意

不被當成是學倫問題 => 學倫在意的是欺騙

- 爛/懶可以被檢驗，騙不易被檢驗，所以不容許欺騙。
- 欺騙才是學倫要處理的問題

抄襲?



https://en.wikipedia.org/wiki/Mona_Lisa

<https://image.posterlounge.ie/images/l/1888787.jpg>

<https://i.pinimg.com/originals/65/73/f2/6573f2c6c387aad57a577e27fc471a85.jpg>

Plagiarism 抄襲

Plagiarism is the “wrongful appropriation” and “stealing and publication” of another author’s “**language**, thoughts, ideas, or **expressions**” and the **representation of them as one’s own original work**. (Wikipedia)

Plagiarize: to steal and pass off (the ideas or words of another) as one's own : use (another's production) without crediting the source (Merriam-Webster Dictionary)

將別人的貢獻讓人誤以為是自己的創見 (竊為己有)

Exceptions: the source is obvious or common knowledge, so there is no mistake that you are taking credit for it.

- To be, or not to be. 床前明月光 $E=mc^2$

抄襲要看的不只是文字的相同，而是意圖或後果

抄襲 ≠ 複製/拷貝

抄襲為何錯誤？有何影響？

- 偷懶 intellectual laziness (對他人不公 Unfair)
- 盜取他人名聲
 - 被盜用者名聲未必受害
 - 名聲非有限資產
 - 可由被抄襲人檢舉
- 影響著作權 copy right，可由出版商或原作者提告 (非學倫)
(Fair use; non-profit use. Copy right protects economic right.)
- Violation of author-journal agreement (journal's decision)
- 剽竊他人研究成果、重複計算研究成果 => 膨風研究成果
=> 影響研究資源 (學位、職位、獎項、計畫) 分配

Sample of an iThenticate report

- 偵測相似度
- 百分比只是參考，可以作為是否人工檢視的門檻
- 重點在於相似文字在文章中的重要性

Bibliography excluded.

iThenticate® Sample Manuscript.docx

Quotes Included
Bibliography Excluded 66%
SIMILAR

Match Overview

Match	Source	Words	Percentage
1	Crossref	40 words	25%
2	Publications	35 words	21%
3	Internet	25 words	15%
4	Crossref	7 words	4%

PAGE: 1 OF 1

Text-Only Report

ORI provides working definition of plagiarism

“As a general working definition, ORI considers plagiarism to include both the **theft or misappropriation** of intellectual property and the **substantial unattributed textual copying** of another's work.”

“Substantial unattributed textual copying of another’s work means the unattributed verbatim or nearly verbatim copying of sentences and paragraphs which materially **mislead the ordinary reader regarding the contributions of the author.**

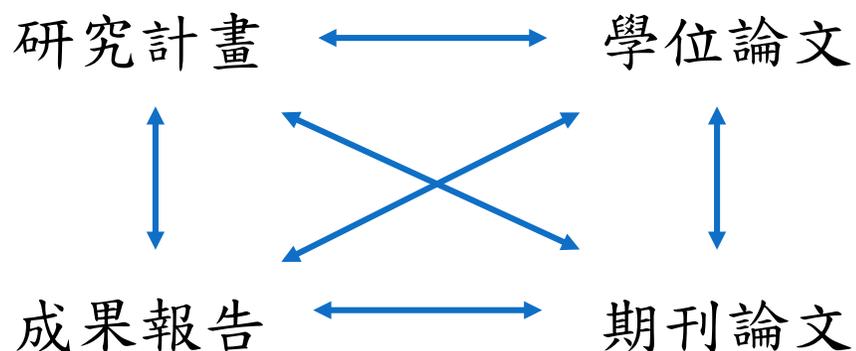
ORI generally does not pursue the limited use of identical or nearly-identical phrases which describe a commonly-used methodology or previous research because ORI does not consider such use as substantially misleading to the reader or of great significance.”

抄襲部分是否為文中的核心？

背景介紹、研究方法，常見取材自他人的論文/計畫，如非涉及創新核心部分，且註明出處(未必是理想引註方式)，則無將他人成果竊為己有之企圖。

- 如果篇幅甚大，且未引註出處，可能視為抄襲。
- 人文社會領域較重視表達，可能視為抄襲

學生/老師/合作者之間關係



- 視為共同著作
- 註明貢獻 (be specific)、註明出處
- 註明於期刊論文及學位論文的 Acknowledgment
- 註明於計畫及成果報告之開始處

無侵占他人貢獻之意圖 => 無抄襲問題

- 應於計畫開始處即註明，文中即無須一一引注。
 - 本計畫依據(衍生自)xxx的碩士論文及本人與xxx合作之論文(2020 PNAS)
- 共同著作(共同發表論文、計畫共同主持人) => 可合理使用
- 並非曾有共同合作，就可以使用對方所有發表文字
- 參與研究團隊討論，而主張為共同創作，除非被引用者提出反對 => 只能接受

避免捲入抄襲爭議

- 投稿前先自我檢驗(文字相似性檢驗軟體)，以避免爭議
- 具體建議國科會：
 - 計畫申請書上加一項：
 - 是否已用文字相似性檢驗軟體（如iThenticate、Turnitin）檢驗

抄襲？

老師去高中作科普演講，用了某論文中的圖表，但未註明出處

- 非學術著作，非學倫問題（著作權是出版商的事）

6. **註明他人的貢獻**：如引用他人資料或論點時，必須尊重智慧財產權，註明出處，避免誤導使人過度認定自己的創見或貢獻。如有相當程度地引用他人著述卻未引註而足以誤導者，將被視為抄襲。此節有以下四點補充：
- 如抄襲部分**非著作中核心部分**，例如背景介紹、一般性的研究方法敘述，或不足以對其原創性構成誤導，應依該領域之慣例判斷其嚴重性。
 - 未遵守學術慣例或不嚴謹之引註，也許是撰寫者草率粗疏，其行為應受學術社群自律（或由本會學術司去函指正），雖不至於需受本部處分，但應極力避免，並應習得正確學術慣例及引註方式。
 - 同一成果如為多人共同研究且共同發表，當然可算做各人的研究成果。如為多人共同研究成果但分別發表（例如同樣調查數據，但以不同方法或角度分析），則**應註明其他人的貢獻**（例如註明調查數據的來源），如未註明則有誤導之嫌。
 - 共同發表之論文、共同申請之研究計畫、整合型計畫總計畫與子計畫，皆可視為共同著作（全部或部分），對共同著作之引用不算抄襲。如依該領域慣例所指導學生論文由老師及學生共同發表，則**指導老師可視為所指導學生論文之共同作者**，但援用時應註明學生之貢獻。

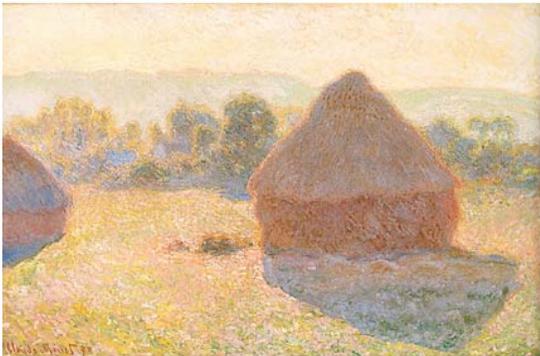
自我抄襲?



1



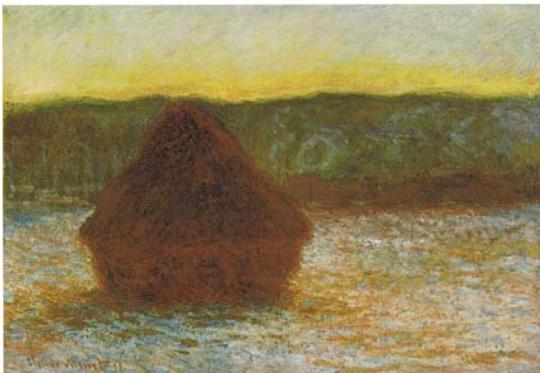
2



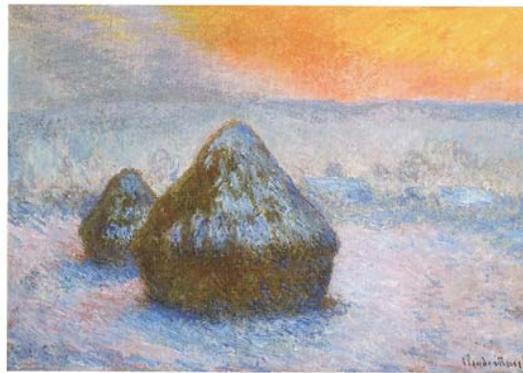
3



4



5



6

Text recycling **within the same paper**

Sentences in the **Summary, Results, Figure Legends** and **Discussion** sections in **the same paper** are highly similar.

Not an issue (so far). Why?

Does not involve duplicate credit.

We don't really care about text recycling.

[自我抄襲]是學倫的假議題

- 抄襲：「竊為己有」（將別人的貢獻讓人誤以為是自己的貢獻）
- How can you steal from yourself?

為何不該自我抄襲？

- 著作權（出版商的事，非學倫）
Protects **economic** rights
- 浪費審查資源（寫個爛計畫？）
- 偷懶、不公平（其他條件的不公平？）
- 重複發表（**duplicate publication**）
 - 重複計算研究成果（重複獲利）、影響研究資源分配
- 真正問題是[重複發表]！

“According to our analysis, the most common uses of text recycling in scientific research articles are **generally allowable** according to U.S. copyright law under **fair use**. Reuse of materials that describe methods, materials, background, and literature review are especially **likely to pose little legal risk.**”

Text Recycling in Research Writing: U.S. Copyright Law and Fair Use. Text Recycling Research Project. White Paper. May 26, 2021

[自我抄襲]應改為[重複發表]

重複發表：

- 涉及論文創新核心內容
- 重複獲取研究成績，重複獲利 => 學倫

文字再使用：

- 不涉及論文創新核心內容
- 不限於文字(如圖表、公式、配方)
- 非學倫問題

[自我抄襲]有負面意涵，被[抄襲]的字面綁架，應避免使用。

D. Juyal, V. Thawani, and S. Thaledi, Plagiarism: An Egregious Form of Misconduct, *N. Am. J. Med. Sci.*, 7(2), 77–80, 2015.

孫以瀚. 論自我抄襲—重複發表、文字再使用，有無學術倫理上的處罰必要？2020.11.15 科技報導

是否重複發表要看是否涉及核心成果

相同作者的兩篇論文

- Introduction/Background (未必是核心)
- Material and Methods (非核心，可引用，但應引註前篇論文)
- Results (核心)
- Discussion (核心)
- References (本非論文中的創作部分)
- 除非相同部分為核心成果，否則不至於重複計算成果
- 領域差異：
 - 實驗科學：容易界定核心成果(Results, Discussions)
 - 人文社會：可能看重寫作表達，認為整篇文字都是核心
- 遵守個別期刊的規範

相同結果不能重複發表

例：論文A為40案例的臨床實驗，論文B增加了20個案例，應說明60個案例中有40個案例已發表於論文A，否則將誤導認為60個為新案例。

兩篇論文核心內容高度重疊，如在後文中未註明部分結果出自前文
=> 重複發表（重複獲利）。

在統合分析（meta-analysis）上，如對藥物療效的統合分析，會造成錯誤的結論，因為相同的研究結果會被重複計算。

重複發表的容許狀況

完全相同內容，以同樣文字或翻譯成不同文字，發表在另一期刊，有註明出自前文，投稿時有告知期刊。

- 為了不同讀者群
- 主編知情且接受
- 在文後(acknowledgment)說明出自前文
- 列在著作目錄，但有說明

內容大幅重複，但不涉及核心成果

藥物測試 (drug testing)

- 不同藥物、相同測試方法、相同結果呈現方式、相同文字、圖表、架構
- 不同結果（數值）
- 並非同一研究結果重複發表

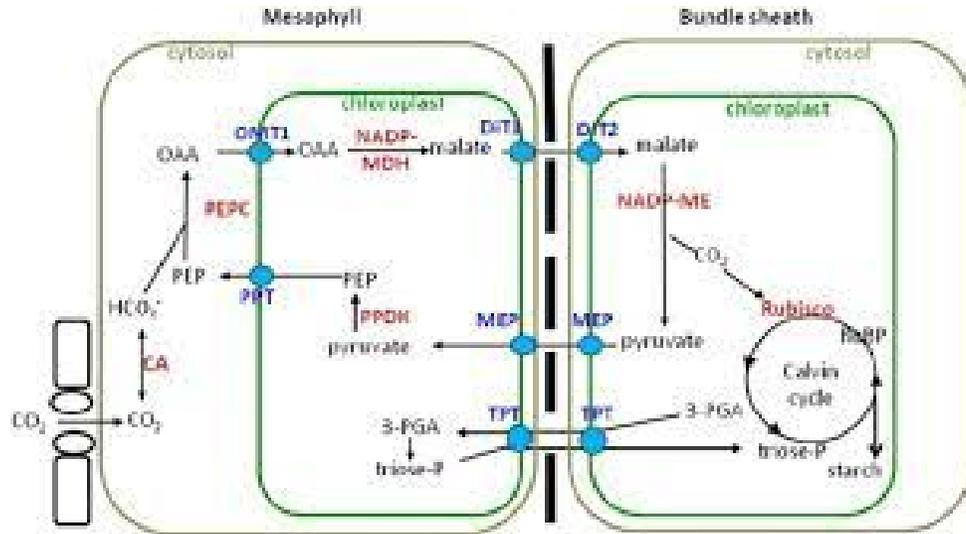
關鍵不在於內容（文字、圖表）的重複，而在於是否會使人以為這是未發表的研究成果，以致重複獲取研究成績（credit）。

Incremental, fragmented publications，是不同的發表策略，不論是否可取，並非罪惡，偶而也有其必要性。

累積100種藥物測試結果再發表—未必更有價值。

背景介紹的重複，不涉及核心成果

Fig. 1. The C4 pathway for carbon fixation



每篇論文都是在既有的基礎上，提出部分的創新。在敘述已有基礎時，即使內容重覆，也不至於讓人誤以為是自己創新的貢獻。

(Taken/modified from Kakri et al. 2013;
with permission from the publisher)

研究方法為了讀者方便，可以重複

Method

The ... is [as previously described \(xxx\)](#), with the following modifications.

“it may be preferable to provide only a brief summary or paraphrase of that material and refer readers to the source document for details .”

“Many editors, however, prefer that articles published in their journal be **self-contained**—especially if the source document is behind a paywall.”

For the convenience of the readers

UNDERSTANDING TEXT RECYCLING: A Guide for Editors

By Susanne Hall, Cary Moskovitz and Michael Pemberton for the [Text Recycling Research Project](#); V.1, June 2021

文字再使用 (Text recycling)

一項針對300多份學術期刊的主編及86個英文學術期刊（含括理工、社會科學、人文和藝術等16個領域）編委的調查發現，83.6%的人認為「文字再使用」在某些狀況下是可以容許的，至於是什麼狀況卻沒共識。贊成者認為，既然花了很多心力精雕細琢描述一個方法、事件或觀念的句子，沒有必要每次都重寫，因此「文字再使用」是為了效率與精確。

如果所用的研究方法跟前篇論文一樣，只在文中註明出處，讀者就必須自己去找原始描述，若某些論文需要付費或訂閱，則會造成讀者的障礙。所以「文字再使用」在某些狀況下可以提昇效率與精確，對作者與讀者都有益。

Online publication has less constraints in journal space.

S. Hall, C. Moskovitz, and M. Pemberton, Attitudes toward text recycling in academic writing across disciplines, *Accountability in Research*, 2018.

孫以瀚. 論自我抄襲—重複發表、文字再使用，有無學術倫理上的處罰必要？2020.11.15 科技報導

文字再使用

為了描述的精準、為了讀者方便(研究方法)，可以再使用自己已發表之文字

背景介紹，研究方法、材料，可以出自自己過去計劃或著作，但仍需引註。

建議於計畫起始處註明[本計畫背景介紹及研究方法部分衍生/改寫自本人100年科技部計畫[xxx]及111年*Nature*論文]

建議於論文Acknowledgment處註明”This work is (partly) derived from XXX’s Ph.D. thesis of XX University, 2022)”

Text recycling

相同研究主題，相同研究方法，稍微改變內容

- 同一基因在**不同條件**下的變化 (mRNA level, protein level, phosphorylation level)
- X-ray 解**不同的蛋白結構**

健保資料庫

- Correlation between X_i and $Y_j \Rightarrow$ many papers (highly similar)

背景介紹、研究方法、數據呈現方式，高度相似

每篇論文都是實際進行了研究，不是以相同研究成果重複發表

Value/significance? **Not ethics problem.**

美國 ORI 的立場

美國聯邦研究誠信辦公室（Office of Research Integrity, ORI）並不將自我抄襲列為不當的研究行為（research misconduct）

國際上的作法

未公開發表之著作（如研究計畫、計畫成果或進度報告、獎項申請文件、研討會壁報、研討會摘要、網路檔案如bioRxiv等），包含國際期刊論文寫作倫理規範龍頭COPE在內的組織，一般也未納入自我抄襲的範疇。

同一研究成果在研究過程中的系列性文件不算自我抄襲

同一研究成果在研究過程中的系列性文件(計畫、IRB申請、計畫成果或進度報告、學生進度報告、學位論文、研討會論文、期刊論文、網路檔案)彼此之間不應視為自我抄襲，因不涉及重複獲利。也無須自我引註。

- 著作目錄上只能列出一項
- 其他(例如論文集)如果列入著作目錄，就該明確註明其性質
- 學位論文在未發表於期刊前，可列為著作。但如果已(部分或全部)發表於期刊，就該註明或不列入。
- 一般期刊不把學位論文、研究計畫、成果報告當成已公開發表。

不同期刊、領域的態度不一

***Anesthesia & Analgesia* Instructions for Authors:**

“Provided the authors are not engaged in duplicate publication, the Journal **does not view “self-plagiarism” as misconduct**. Authors are permitted to reuse their own words, and are encouraged to do so when describing identical research methods in multiple papers.”

Many journals do not allow self plagiarism or text recycling.

In **humanities and literature**: “novelty and the essence of the work are in the eloquence and the wording,”

Habibzadeh and Marcovitch (2011) Plagiarism: the emperor’s new clothes. *European Sci Editing* 37:67-70.

Follow the current rules of journal and funding agency.

Don't be controlled by our tools

Tools change our behavior

Homology-detecting software (e.g. iThenticate) => catch plagiarism!

=> self-plagiarism (not because we care, but because we can detect it.)

路燈下找鑰匙的醉漢：鑰匙沒掉在這，但這裡比較亮

Beware of conflict of interests in opinions



telephone +1 (510) 764-7800

email INFO@ITHENTICATE.COM

web WWW.ITHENTICATE.COM



WHITE PAPER

THE ETHICS OF SELF-PLAGIARISM

Turnitin to Be Acquired by Advance Publications for \$1.75B

By [Sydney Johnson](#) Mar 6, 2019

<https://www.edsurge.com/news/2019-03-06-turnitin-to-be-acquired-by-advance-publications-for-1-75b>

“While iThenticate’s paper may be perceived as an authoritative guide, it is also the product of a **for-profit** plagiarism-detection business with a virtual monopoly across academe, scholarly publishing and government.”

“a **misrepresentation** of the realities and ethics of academic research - and a guide that **leads to worse, not better, writing.**”

Cary Moskowitz and Aaron Colton (2021) Resources on avoiding self-plagiarism are scarce and problematic (opinion). *Inside Higher Ed.* <https://www.insidehighered.com/print/views/2021/03/05/resources-avoiding-self-plagiarism-are-scarce-and-problematic-opinion>

7. **自我抄襲的制約**：研究計畫或論文均不應抄襲自己已發表之著作。研究計畫中不應將已發表之成果當作將要進行之研究。論文中不應隱瞞自己曾發表之相似研究成果，而誤導審查人對其貢獻與創見之判斷。自我抄襲是否嚴重，應視抄襲內容是否為著作中創新核心部分，亦即是否有誤導誇大創新貢獻之嫌而定。此節亦有以下兩點補充：
- a. 某些著作應視為同一件（例如研討會論文或計畫成果報告於日後在期刊發表），不應視為抄襲。計畫、成果報告通常不被視為正式發表，亦**無自我引註之需要**。研討會報告如於該領域不被視為正式發表，亦無自我引註之必要。
 - b. 同一研究成果以不同語文發表，依領域特性或可解釋為針對不同讀者群而寫，但**後發表之論文應註明前文**。如未註明前文，且均列於著作目錄，即顯易誤導為兩篇獨立之研究成果，使研究成果重複計算，應予避免，但此應屬學術自律範圍。

三、(研究人員違反學術倫理之行為類型)

(三)抄襲：援用他人之申請資料、研究資料或研究成果未註明出處。註明出處不當情節重大者，以抄襲論。

(四)自我抄襲：研究計畫或論文未適當引註自己已發表之著作。

(五)重複發表：重複發表而未經註明。

我的建議：

- 抄襲：抄錄他人作品，未註明出處，意圖誤導為自己的創見或成果，情節重大者。
- 重複發表：將相同研究成果重複發表，卻未適當引註，以致有重複獲取研究成績之虞，且情節重大者。（計畫書、成果報告不算發表）
- 取消「自我抄襲」項目。

計畫的自我抄襲?

非重複發表

同一計畫內容向校內申請獲小額補助，再向國科會提出申請

- 部份內容可以重覆，但要執行的部分不應重複
- 可以說明雖獲補助，但經費不足，需另申請經費始能完整執行
- 在計畫明顯處(摘要、第一段說明、粗體字)說明

多年期計畫申請每年300萬，僅獲國科會補助一年，且只有100萬，第二年再以同樣內容向國科會提出申請。

- 可以說明僅獲補助，且經費不足，因此再度申請，以完成原本申請內容
- 針對去年審查意見逐一回覆，並說明這一年的進度及計畫內容的修改

去年向國科會申請計畫未通過，今年以相同內容的計畫再度申請

- 針對去年審查意見逐一回覆，並說明這一年的進度及計畫內容的修改

計畫的自我抄襲？

非重複發表

送交國科會的研究計畫，其「過去研究成果」或「先期研究成果」大幅引用自己已發表之期刊論文

- 「先期研究成果」(Preliminary Results)不是unpublished results，是本計畫的基礎，用以顯示自己在此題目上已有一定基礎，可以是申請人**已發表**或**未發表**的成果。
- 如為已發表的結果，可以引註出處，可以增強其可信度(已經過peer review)，但未引註也非錯誤，因為並不涉及重複發表及獲利。

學位論文後續發表於期刊

學位論文各章節改寫(或未改寫)後續發表於期刊

明確告知

- 建議於期刊論文Acknowledgment 中說明(This work is based on (or modified from) Chapter 2 of my Ph.D. thesis in XX University, 2020)
- 投稿時也於cover letter中告知 editor

不要有重複獲利的意圖

- 如學位論文尚未於期刊發表，可以當成著作。一旦以相同內容發表於期刊，就不該列入著作目錄。
- 如學位論文的部分章節改寫後發表於期刊，應當於著作目錄中註明。

國科會對一稿二投的規範

國科會對研究人員學術倫理規範第八點（108年11月21日修正）

8. **同一研究計畫不得重複申請補助**；論文一稿多投應遵守發表單位（期刊與會議等）之出版倫理規定：

(1) **同一研究計畫不得同時重複向本部提出申請**。以同一研究計畫向本部及其他機構申請補助時，應於計畫申請書內詳列申請本部及其他機構補助之項目及金額，**同一項目及金額不得重複申請補助**。

(2) 論文是否被允許一稿多投，應遵守發表單位（期刊與會議等）之出版倫理相關規定。

國科會補助專題研究計畫作業要點

第二十六點(略以)：

(五) **同一研究計畫不得同時重複向本部提出申請**，違反規定者，**依本部學術倫理案件處理及審議要點規定處理**。

(六) 以同一研究計畫向本部及其他機構申請補助時，應於計畫申請書內詳列申請本部及其他機構補助之項目及金額，**同一項目及金額不得重複申請補助**。

一稿二投、一魚兩吃

系列性研究難免主題相似，申請人有責任明確說明此計畫與其他執行或申請中計畫的相關性與差異性，以避免引起誤解及爭議。

建議說明：本計畫申請內容與本人執行中另一計畫內容部分重複，因執行中計畫僅獲補助100萬元，只夠執行原計畫中1/3樣本的分析，因此本計畫申請補助以執行另外2/3樣本的分析，樣本不同，經費並非重複使用。

科技部生科司陳司長敬致學界信函 (110.8.20)

今年科技部首次建置「專題研究計畫申請書比對系統」，目前系統以110年度大批專題研究計畫申請書作為data base進行測試，比對結果顯示本司研究計畫內容CM03，相似度區間在60%以上者計有40對資料(亦即涉及80個計畫)。本司進一步以人工檢視這40對資料，發現其中有些明顯為一稿多投，同一計畫以相同或不同申請人申請同一學門、跨學門或跨司計畫。另外也發現，有些計畫(不同申請人)會使用相同的文章模板(template)，依不同研究標的(target)些微改寫或增修背景、策略、執行方法等部分；或者不同計畫之研究主軸或對象雖不同，但卻有極為雷同的實驗設計與方法等情事。上述情節嚴重者，本司將會召開專案審議會議，討論是否涉及違反學術倫理。

為避免上述情形，特別在此提醒大家，撰寫計畫書時應當在計畫書明顯之處註明應當揭露的事實，如計畫申請書部分內容如屬學生之學位論文，應於計畫書內容揭露及引註；部分內容及圖表如有引用自己或他人已發表之論文，應確實註明出處；延續性或多年期計畫應揭露之前計畫並說明研究之差異性；計畫互為共同合作應有分工說明；不得將他人已完成之研究成果，作為自己計畫申請及研究成果等。

建議計畫主持人於計畫申請案送出前，自行完成「原創性比對」相關資料。可先利用比對軟體確認是否有引用不當之疑慮，及時修正計畫文稿內容，不僅可提升計畫品質，亦能降低學術倫理風險。

生成式 AI

A generative artificial intelligence or generative AI is a type of AI system capable of **generating text, images, or other media** in response to prompts. Generative AI systems use generative models such as **large language models** to statistically sample new data based on the training data set that was used to create them.

(Wikipedia)

生成式人工智能或生成式 AI 是一種能夠**根據提示**生成文本、圖像或其他媒體的 AI 系統。生成式 AI 系統使用生成模型（例如**大型語言模型**）根據用於創建它們的訓練數據集對新數據進行統計採樣。（by Google 翻譯）

依據大數據資料庫，透過人工智慧學習，生成自然語言

- 要有大數據資料 => 不同文字的成熟度會不同
- 可以透過交談不斷學習
- Misinformation and bias
- 生成式，並非copy

ChatGPT 人工智慧聊天機器人程式



Chat **G**enerative **P**re-trained **T**ransformer

Menu

<https://openai.com/blog/chatgpt>

Introducing ChatGPT

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.

[Try ChatGPT ↗](#)

[Read about ChatGPT Plus](#)



AI要驗證你是人類！

進入 ChatGPT 網頁



驗證您是人類



Query



Examples

"Explain quantum computing in simple terms" →

"Got any creative ideas for a 10 year old's birthday?" →

"How do I make an HTTP request in Javascript?" →



Capabilities

Remembers what user said earlier in the conversation

Allows user to provide follow-up corrections

Trained to decline inappropriate requests



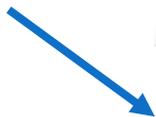
Limitations

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021

寫入問題



Query and Response

YI

我發現在TNBC細胞中，STAT磷酸化增加，要證明STAT磷酸化與癌化的關係，需要做哪些實驗？



如果您想進一步證明STAT磷酸化與TNBC的癌化有關，以下是一些可能有幫助的實驗：

1. 確定STAT的磷酸化位點：在TNBC細胞中，磷酸化的STAT可能出現在不同的位點，進一步確定磷酸化的位點可以幫助證明STAT磷酸化與TNBC癌化的關係。這可以通過質譜分析和蛋白質修飾實驗等方法進行。
2. 研究磷酸化STAT的影響：研究磷酸化STAT對TNBC細胞增殖、侵襲和轉移的影響可以進一步確定磷酸化STAT與TNBC癌化的關係。可以通過使用磷酸化缺陷的STAT蛋白來進行此類實驗。

↻ Regenerate response



搜尋網路，會有很多類似的句子，但可能沒有完全一樣的句子。
(作者得負責確認沒有抄襲)

從大數據資料庫學習而來，但非直接引用，也無從知道資訊的來源。

Regenerate response

- 每次答案均不同
- 即使兩人問同樣問題，答案也不完全一樣。

Send a message...



有辦法分辨AI與人類的作品嗎？

無！

- 分辨工具發展中，只能推測，無法證明
- 我不預期能發展有效可靠的辨別工具

ChatGPT 目前的問題

- 學習資料庫只到2021
- 沒有引注資料來源
- 資料不可靠，會捏造 => 需要人工檢驗
- 可能被誤導 (training set, algorithm) (政治力、現存歧視偏差、假資訊)
- AI 運作的過程是黑盒子
- 擔心洩密 (提問的資訊進入學習資料庫)
- 數據圖表分析
- 被主流意見主宰，少數意見被壓制

這些問題在AI出現之前都存在!

- 可預期很快會改進
- 很多競爭者

ChatGPT 官方在 5 月 12 號釋出 Web browsing 與 Plugins 的功能，目前僅提供付費版使用者搶先體驗

未來會更強大

BERT: Bidirectional Encoder Representations from Transformers (Google)

RoBERTa: Robustly Optimized BERT Pretraining Approach (Facebook)

ChatGPT for Google

GPT-4 (OpenAI)

文心一言 Enhanced Representation through Knowledge Integration (ERNIE Bot) (百度)

New Bing: GPT4 + 搜尋

ChatPDF <https://www.chatpdf.com/>

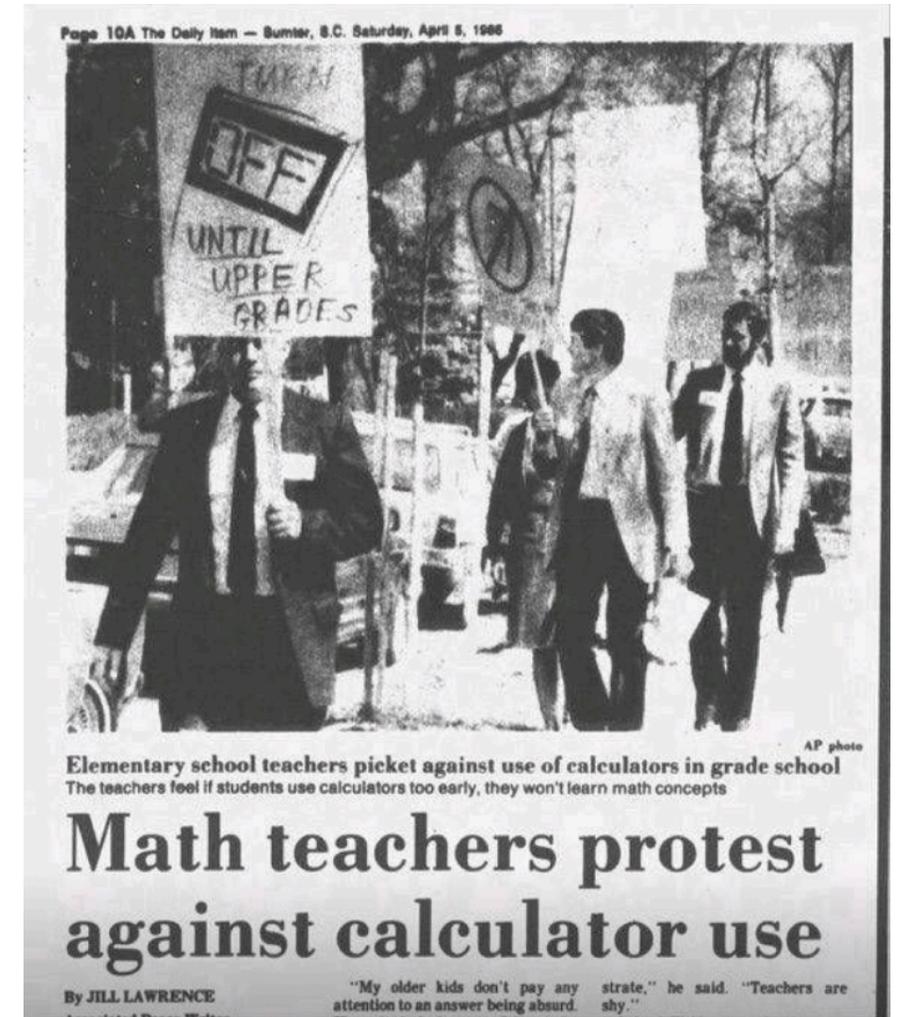
Domain-specific

- BioGPT
- BioBERT and PubMedBERT

TOME: 做簡報

AI 勢不可擋

- 革命性、強大的新工具!
- 破壞性創新!
- 利遠大於弊，勢不可擋!
- 阻擋、禁止，是無效的!
- 不擅用新工具，就落伍!
- 勢必改變人的行為!
- **人必須調整以適應及利用AI!**



<https://www.linkedin.com/pulse/how-adapt-knowledge-testing-professional-training-ai-world-espanet/>

破壞性創新 Destructive Innovations

蒸汽機、電、交通（輪子、火車、飛機、自駕車）、
個人電腦、Word Processing（中文打字、修改、語言轉換）、手機、網路、
生物醫學（recombinant DNA, PCR, CRISPR/Cas9）

Google搜尋引擎、Google Map

ChatGPT

- Free and open for all. (免費版) => 有利平等，提高baseline
- 會使用新工具者佔優勢
- 付費（功能升級）=> 不平等
 - 永遠有不平等

規範、管理? Governance?

AI developers

- Tool developers always act/react faster than government.
- Community self-imposed regulation/guidelines? (Recombinant DNA, 風險分級管控)

Users

- Impossible to regulate
- Selection by **market**
- 規範的理由：洩密（政府、企業）
 - 審查人對審查資訊應保密（是否視為洩密？非公開、非給特定對象）

歐盟 AI 法案

2023.6.14

歐盟《AI法案》草案大綱

- 1 禁止業者在公共場所設置感應裝置收集生物辨識資料
- 2 禁止業者收集監視器或網路影像來建立臉部辨識資料庫
- 3 禁止警察單位利用過往犯罪紀錄來發展犯罪預測系統
- 4 業者在設計AI模型時須預防AI模型創造非法內容
- 5 業者須公開AI模型訓練資料中是否含有著作權保護資料
- 6 業者在特定情況下須公開創作內容來自人類或AI模型
- 7 違法者罰金最高可達全球年營收的6%至7%

資料來源：《華爾街日報》

製表：陳穎芃

- 規範業者、執法單位
- 風險分級

在研究上的應用

- 改英文、抓重點、改寫、摘要、縮短、寫論文
- 整理背景資料、文獻
- 改程式、寫程式
- 瞭解過去，整理總結，避免重複，避免錯誤，促進嚴謹
- 減少投入的時間心力，方便
- 來自過去群體經驗，**趨向平庸（四平八穩、面面俱到、中規中矩、mediocrity）**
- **無法創新？（建立不同領域之間的連結）**
- 找出重要未解問題
- 設計實驗
- 寫計畫、寫論文
- 審查計畫、審查論文

在研究上的應用

研究者：

- 這個領域有什麼未解決的重要問題？
- 這個領域有些什麼已知的研究？有些什麼相關文獻？
- 我的研究結果可以得出什麼結論？有什麼創新點？
- 對照組是否足夠？還需要補什麼實驗？
- 未來可以有什麼延伸的研究方向？
- 可以投什麼期刊？

期刊、資助單位：

- 這個研究有些什麼創新之處？重要性？應用價值？
- 有哪些類似的研究？是否能超越類似的研究？
- 這個研究的結果是否足夠支持其結論？還需要補些什麼實驗？
- 是否符合本期刊的方向？(是否鎖死期刊發展方向?)

Q: 請提供相關的文獻

ChatGPT: 以下是一些關於三陰性乳癌和EGFR的研究文獻，供您參考：

Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, Pietenpol JA. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest.* 2011;121(7):2750-2767. doi:10.1172/JCI45014

Ueno NT, Zhang D, Targeting EGFR in Triple Negative Breast Cancer. *J Cancer Sci Ther* 2014; 6(6): 236-243. doi: 10.4172/1948-5956.1000262

錯誤

Baselga J, Albanell J. Mechanism of action of anti-HER2 monoclonal antibodies. *Ann Oncol.* 2001;12 Suppl 1:S35-S41. doi:10.1023/a:1011199018579

Yardley DA, Kaufman PA, Brufsky A, et al. Treatment patterns and clinical outcomes for patients with de novo versus recurrent HER2-positive metastatic breast cancer. *Breast Cancer Res Treat.* 2021;186(1):107-117. doi:10.1007/s10549-021-06105-5

Hynes NE, MacDonald G. ErbB receptors and signaling pathways in cancer. *Curr Opin Cell Biol.* 2009;21(2):177-184. doi:10.1016/j.ceb.2008.12.010

Moulder SL, Yakes FM, Muthuswamy SK, Bianco R, Simpson JF, Arteaga CL. Epidermal growth factor receptor (HER1) tyrosine kinase inhibitor ZD1839 (Iressa) inhibits HER2/neu (erbB2)-overexpressing breast cancer cells in vitro and in vivo. *Cancer Res.* 2001;61(24):8887-8895.

Siddiqui S, Chopra R. EGFR: A Potential Target for the Treatment of Triple-Negative Breast Cancer. *Chemotherapy.* 2017;62(3):177-184. doi:10.1159/000452984

捏造

我們要如何適應？

- 需要人工檢驗正確性（但人有惰性）
- 責任（在人）（e.g. 醫囑）
- 表達能力（閱讀、搜尋、整理、歸納，語言、文字）降低
- 當AI幫我們做了這些，人要如何展現自己的價值？聚焦於創新？
- 人文社會領域強調表達，未來如何彰顯創新？
- 如何證明是自己原創？
- 過去的評鑑方式必需改變

適應進行中！
（全世界、各行各業）

AI輔助產生的作品，是否視為作者的原創？

作者的貢獻

- 工具是在人的指導下運作
- 即使部分或全部內容由AI生成，**原始問題仍是由作者提出**，並需要對內容正確性、架構、邏輯、取捨做判斷
- 助理幫忙查資料、做實驗、整理數據，但未必列名為作者
- 負責的是作者，非AI

如何評鑑學術成就？

- 期刊或資助機構應該判斷的是**作品本身**是否有發表/資助的價值，是否可靠，而非credit歸屬(除非有人控告抄襲)
- 給獎、聘任，才是看**個人**的貢獻

研究的創新?

我們的研究，有多少創新?

- 依循常用研究思維、研究方法、分析方法（標準套路）
- 研究題目，常追隨別人開創的思路，跟隨潮流
- 換個關鍵詞
- 真正開創的有多少成分?

利用AI可以幫我們找出**常見的研究套路**，常用的研究策略，該做的實驗，該有的controls，找出reviewers可能問的問題，避免犯錯、缺漏。基本上走一條**穩健**的路。

作品只要有部分創新(也許只是結果，未必是研究策略或設計)，也許即有其價值。

作品有無創新，跟創新的來源，是兩回事

- 歷史上，創意的來源經常有爭議，並無損作品本身之創新性

學術成果的評量不該受使用工具影響

真正有爭議的是AI輔助研究所得之榮耀是否歸屬於作者

- 工具的使用，不該是減分，而是**加分**
- 該看的是**成果**，而非花了多少心力
- 整合(人、工具、數據)是一種重要能力
- 研究將多種實驗外包，已屬常態。
 - 實驗A, B, C 分由三位合作者執行，或付費外包給三家廠商執行(非共同作者)
 - 統計、大數據分析，分由兩位合作者執行
 - **重點在於發想、設計、整合**

應評斷研究人員之成果是否有價值，而非其使用工具

對人文領域影響較大

學術界的態度

學術期刊的態度 (e.g. *Nature*)

- 不能列為作者：無法負責 (accountable)
- 使用 ChatGPT 必須註明(於研究方法或致謝中)
 - 當作是工具 (例如：本抗體委託XX公司製備)
 - 使用 ChatGPT 必須註明(於研究方法或致謝中)

Nature: **open and transparency**

AI: lacks transparency (in training dataset, algorithm)

- Hold on to **human verification**
- Develop rules for accountability
- Invest in truly open LLMs
- Embrace the benefits of AI
- Widen the debates

Van Dis et al. (2023) *Nature* 614:224-226

如何註記使用 ChatGPT?

引用 ChatGPT 生成的文字

“As an artificial intelligence language model, ChatGPT does not have legal ownership or copyright over the generated text or figures it produces.” (ChatGPT, personal communication, April 21, 2023) 或

¹ Text generated by ChatGPT, April 21, 2023, OpenAI, <https://chat.openai.com/chat>.

² ChatGPT.

如何註明使用 ChatGPT 幫忙，仍無範例

- The writing of this paper was helped by ChatGPT (OpenAI). (很難註明使用程度)
- The Introduction section was **written/revised** by ChatGPT (OpenAI).
- The original idea was conceived by XX. The literature search/review was done by XX with help from ChatGPT.

註記使用 ChatGPT，是個無法執行的規範

- 無法有效可靠的偵測 ChatGPT 的使用。
- 揭露使用，最多也只說明我用了這工具，用到什麼程度很難說清楚。
- 無法有效執行的規範，是無意義的規範，只會造成所有人的困擾。
- 乾脆不要要求揭露、註記。
- 瓢竊 idea 類似文字抄襲，但難舉證，一般無法處理，也就乾脆不談。這是一樣的道理。
- 把 ChatGPT 視為基本工具，本無須註記 EXCEL、WORD 的使用。

ChatGPT的使用與學術倫理

不當研究行為(Research Misconducts)

- 造假、變造: 作者應對研究成果之正確與嚴謹負責
- 抄襲: 生成式，非抄襲

ChatGPT 的使用

1. 不涉及抄襲
2. 內容的正確性，需人為檢視確認，為作者之責任
3. 創新及價值，本非學倫問題
4. 期刊要求: 作為工具的使用，應揭露註記 (無效的要求)
5. 不應做為審查工具(對審查資訊之保密)

AI生成物的著作權？

AI生成的文字、圖像或程式是否也能歸納為創作？受著作權保護？著作權歸屬？

OpenAI 公司？ ChatGPT程式？ 寫ChatGPT程式的人？向ChatGPT下指令的人？

ChatGPT: AI生成的文字的著作權歸屬問題目前還存在爭議。由於AI本身並非一個具有法律主體資格的實體，因此無法享有著作權。

用AI輔助產生的作品，作者可能被訴沒有著作權？

- 被別人使用

作者的責任

稿費如何計算？

寫手

冠名作者概括承受權利與責任，輔助工具(或人)只對作者負責

確保不引起爭議

Q: 有哪些作品與本作品很類似？

未來的影響

正面影響

- 論文及計畫的文字流暢度、文法正確性，可能提升
- 沒有抄襲的必要
- AI可以提醒避免犯基本錯誤
- 研究更要問對問題，才有創新
- 審查人要更重視判斷作品的創新性

負面影響

- 論文或計畫更容易產出，增加審查負擔
- 內容未必可靠，需要人工檢驗判斷

誌謝

本文中的內容，部分來自2023/3/30臺灣學術倫理教育資源中心AI線上講座觀眾所提出之問題，部分來自2023/6/19中研院[思想櫥窗－AI與未來社會：知識協力、智慧指引？]討論會，謹此誌謝！

感謝聆聽，敬請指教

