AI浪潮下的學術圖書館

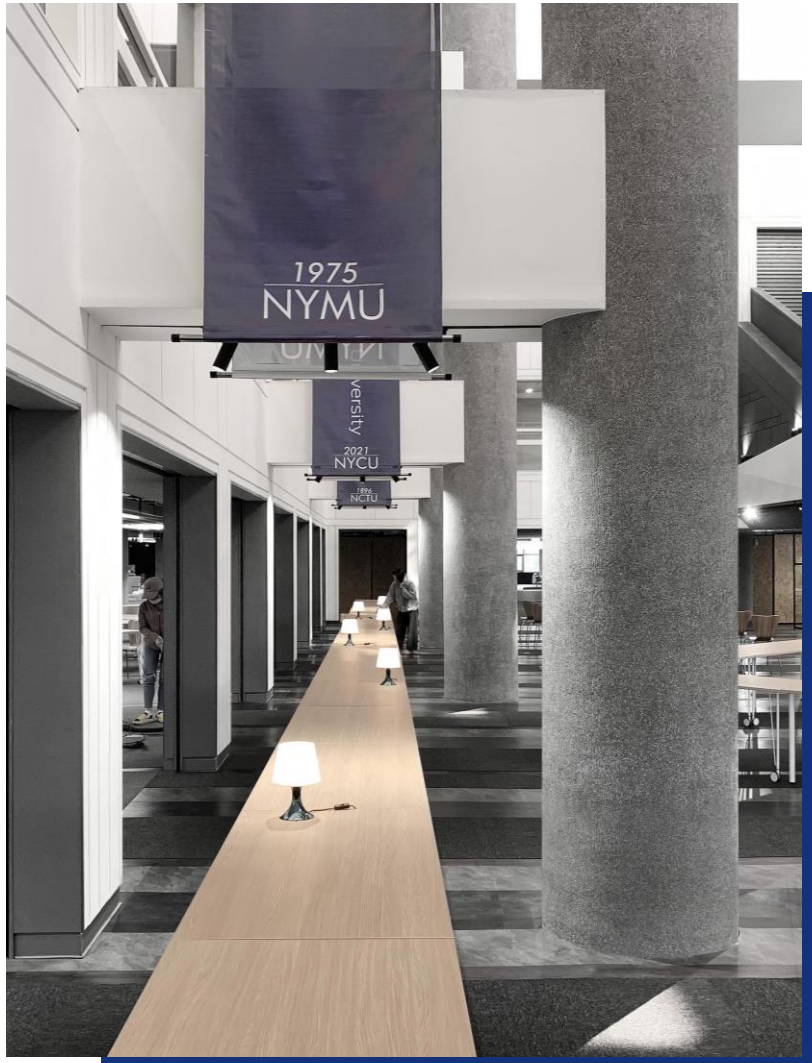
黃明居｜教授，圖書館館長

112. 09. 21

國立陽明交通大學圖書館
National Yang Ming Chiao Tung University Library

國立陽明交通大學圖書館
National Yang Ming Chiao Tung University Library

目錄



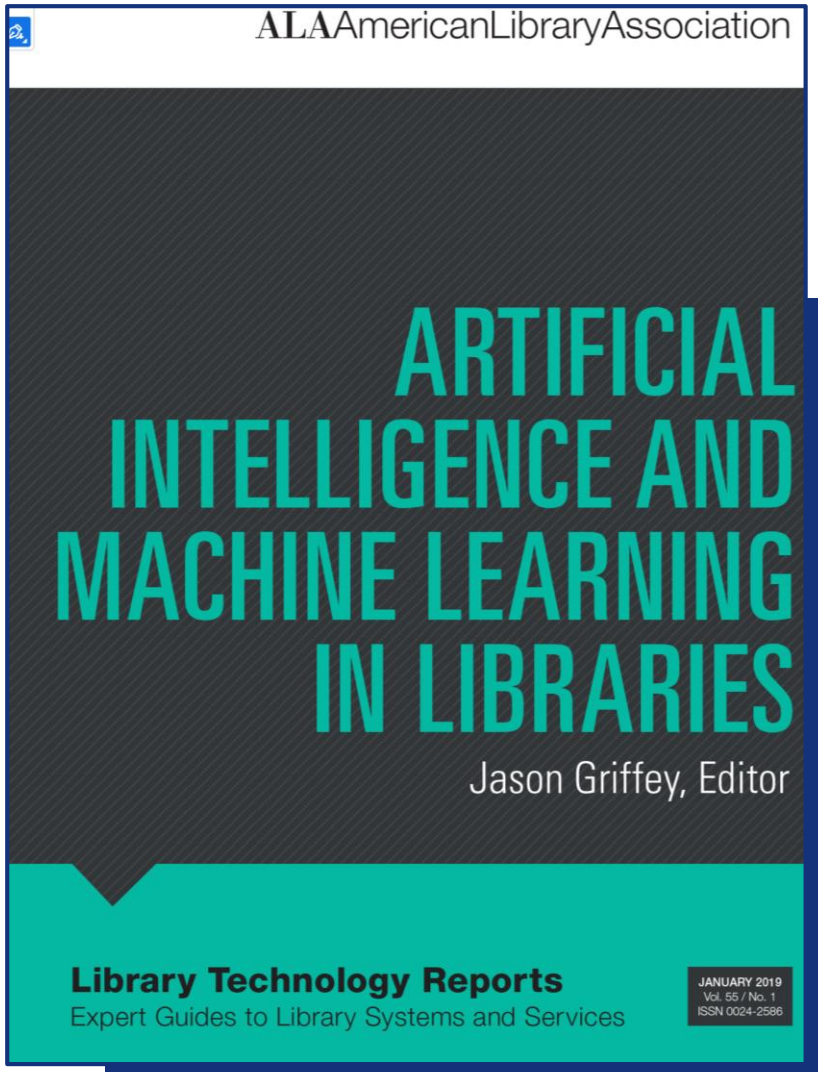
AI時代的大問題

- 當機器能讀取~~（讀懂？）~~圖書館所有資料時，圖書館和館員會發生什麼事（影響）？

  “What Happens to Libraries and Librarians When Machines Can Read All the Books?”（Chris Bourg, 2017）

- 圖書館可以提供怎樣的資訊服務？（Smart Library Services?）

- 圖書館與圖書館館員應如何因應？



AI and ML in Libraries (2019)

- ✓ Introduction
- ✓ HAMLET
- ✓ AI and Creating the First Multidisciplinary AI Lab
- ✓ An Exploration of Machine Learning

Introduction

- in the 1940s, people began to speculate what it would mean for a computer to be “intelligent”
- AI 從符號人工智慧轉向機器學習
  - **人工定義規則 vs.資料產生規則**
  - Ex: AlphaGo (2016)
    - **透過人工神經網路、深度學習進行訓練**
- AI & ML成為現代科技體驗的一部分
  - 人們常常沒有意識到他們正在體驗的是機器學習系統
- 人工智慧優缺點
  - AI的好壞取決於它的訓練數據以及在學習決策時賦予系統的權重
  - 任務自動化、提高效率 (Homecourt) HC2
  - 道德、偏見、社會影響

For a given **neural net,** and a **given training set**, and **a given query**, one could build a statistical model of the likelihood of outcomes, but **not predict with certainty what that outcome might be**. This means that **when biases are present in training data**, the effects they might have on queries and outcomes **may not be directly predictable.**

HAMLET

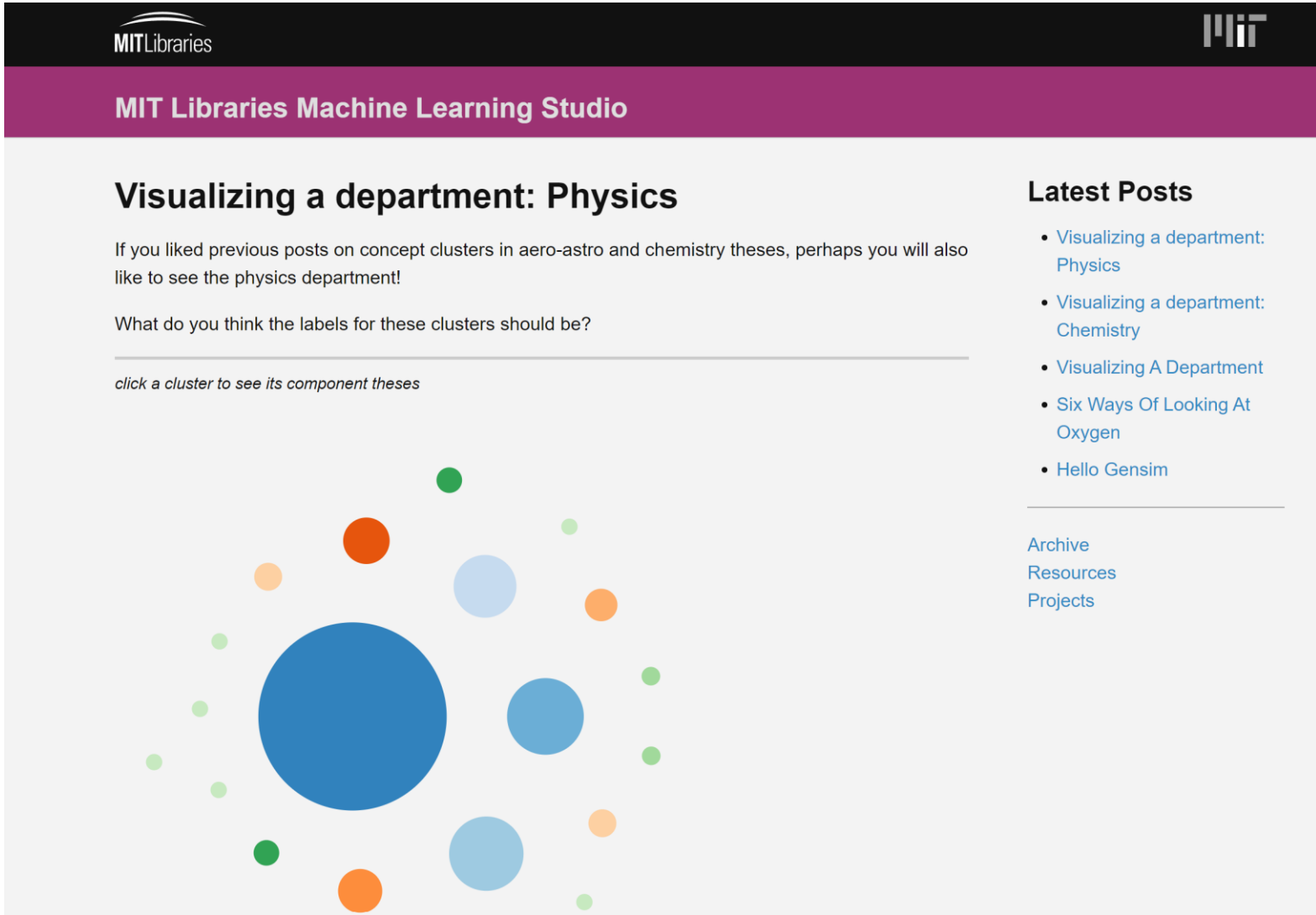
How About Machine Learning Enhancing Theses? - a pilot discovery project (MIT, 2017)

https://hamlet.andromedayelton.com/about/

- 透過機器學習(Word2vec)加強圖書館探索服務
  - 透過 MIT 研究生論文進行訓練，用於分析比較文字
  - 允許使用者上傳文件並接收類似論文推薦
  - 透過相關論文建議引用文章來源，協助進行文獻回顧
- 受限於訓練所用資訊
  - 主要來源：MIT 的 STEM 論文
  - 可能不適合其他研究領域，如藝術史或舞蹈
  - 透過相似度與概率分析，準確度或相關性可能有疑慮
- 使用者仍應意識到工具的侷限性，多元參考其他研究方法

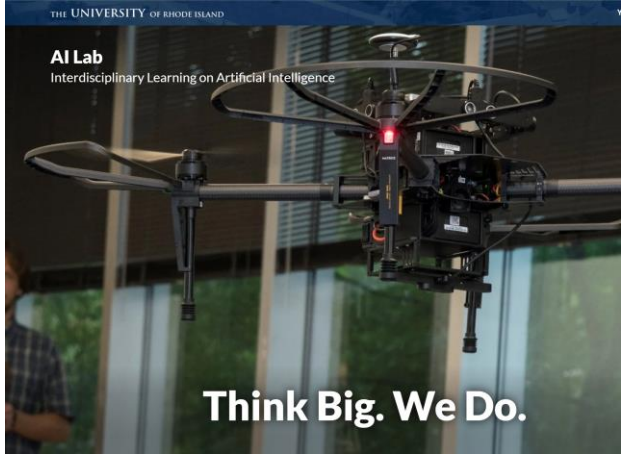
MIT Libraries MLS

How About Machine Learning Enhancing Theses? - a pilot discovery project (MIT, 2017)



AI and Creating the First Multidisciplinary AI Lab

AI Lab at the Univ. of Rhode Island



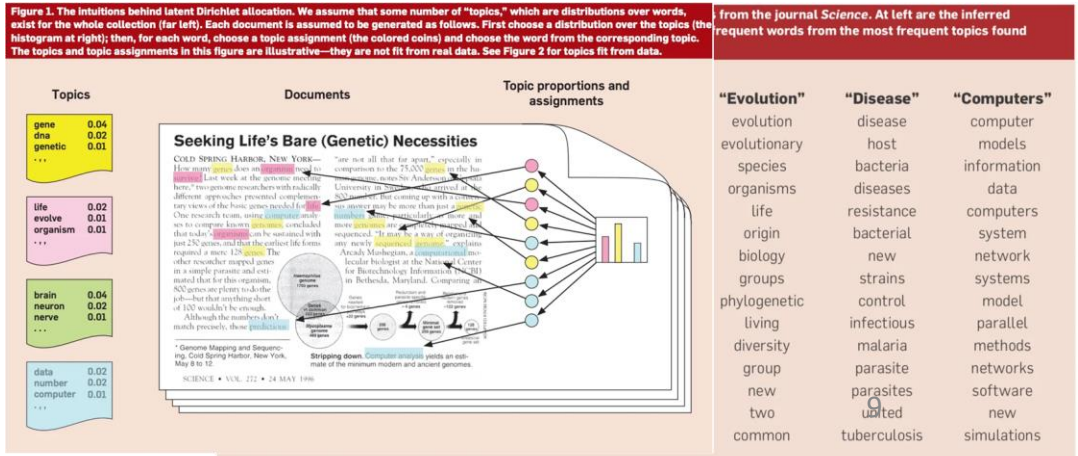
- 核心任務
  - AI 功能教育
  - 促進 AI 研究跨學科合作
  - 促進 AI 社會影響的積極討論
- 推出意義重大
  - 透過產生研究人員間協同作用，讓學校站在發展前沿
  - 與各學科互動合作，讓學生實際使用深度學習演算法來增強穿戴式裝置功能，深入了解如何利用腦電波控制機器人並參與人機關係相關議題討論
- 成為想法與靈感來源、AI 新課程產生器
  - 透過活動提高大眾對 AI 的認識，促進教育推廣
- 透過圖書館的研究中心角色，服務使用者
  - 圖書館的中立性、可訪問性

➢ The idea is similar to that of a makerspace in the library

An Exploration of Machine Learning

Craig Boman

- 研究目標
  - 探索 LDA技術，產生圖書館主題標目（Library Subject Headings (LSH)）
  - Latent Dirichlet Allocation
- 資料處理
  - 以 Gutenberg 計畫電子書為基礎
  - 透過 bash 命令檢索、提取電子書資料
  - 透過 bash 和 Python 進行資料轉換與載入
- 透過統計學原理延伸的機器學習預測



AI and ML in Libraries

- 電子書籍或期刊的提供商(擁有大量數字化文本語料庫)
  - ✓ 嘗試以AI和ML為基礎的**新索引與搜尋服務**
- 機器學習系統經過訓練，創建Metadata的潛力非常高（**人工智慧的編目系統**）
  - ✓ 更加注重培訓數據的準備和產出的評估，而不是直接創建描述（**館員的重要角色**）
- **個人化服務**：隨著系統根據讀者的行為進行自我訓練，隨時間的推移，系統會繼續學習
- 研究人員和學生將擁有AI系統，協助他們尋找資訊、總結資訊並建立個人參考書目等
- 如何對待這些系統的**知識產權**將對圖書館未來如何使用、蒐集、共享和保存等將會有長期影響
- AI&ML系統值得圖書館和圖書館員密切關注!
- 當機器人能夠寫出與人類所寫的論文毫無區別時，**教育將如何改變？**

AI對圖書館的影響

- ✓ How AI can improve information organization, accessibility, user services, and library analytics.
- ✓ It also emphasizes the importance of AI literacy for both librarians and patrons in today's society.

Source: https://www.aje.com/arc/ways-artificial-intelligence-impacts-libraries/

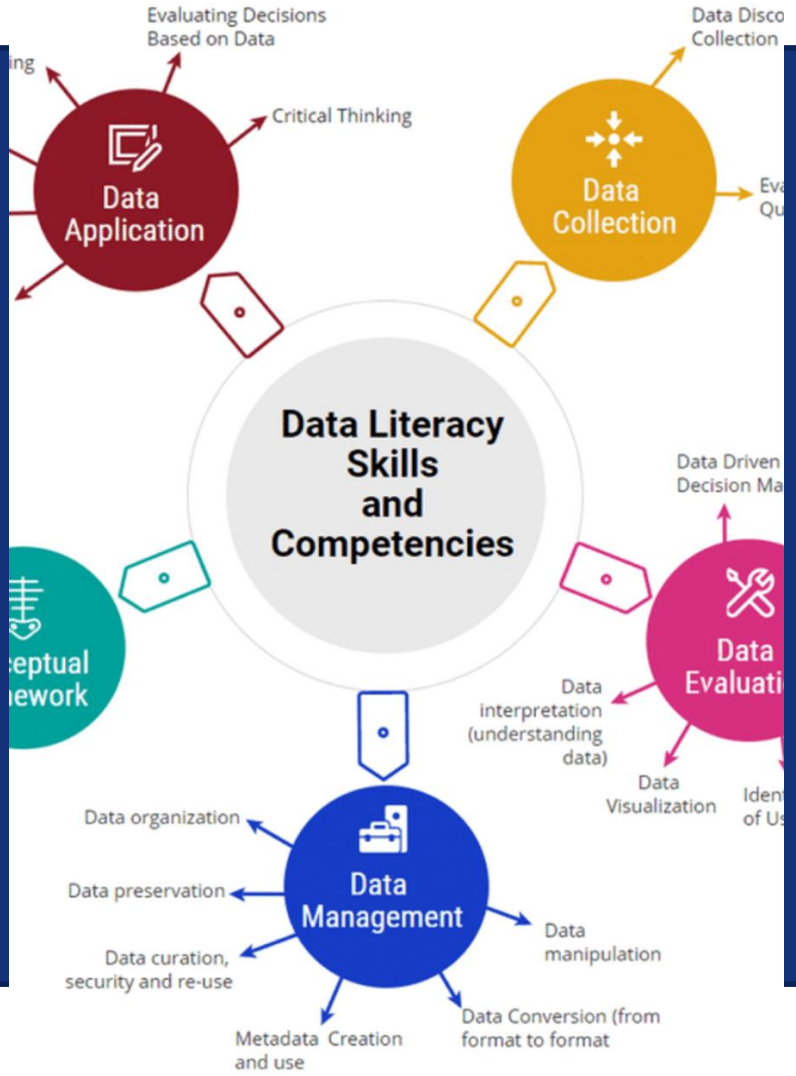
AI對圖書館的影響（續1）

Information professionals

- ✓ librarians are improving the precision of search and recall efforts.
- ✓ librarians will to educate the public on how to locate and interact with these AI tools.

Library operations

- Robotic process automation
- Smart Libraries



Data and AI literacy

- ✓ Libraries and library professionals takes aim at data literacy and AI literacy.
- ✓ Data literacy deals with learning how to locate, understand, and think critically about data
- ✓ AI literacy entails an understanding of its function, logic and limitations, and potential impacts.

https://www.aje.com/arc/ways-artificial-intelligence-impacts-libraries/

AI對圖書館的影響（續2）

- **Library analytics**
- ✓ AI can be used to analyze library data to identify patterns and trends.
- ✓ This information can be used to improve library services and make better decisions about collection development and staffing.
- ✓ **Ex:?**

So https://www.aje.com/arc/ways-artificial-intelligence-impacts-libraries/

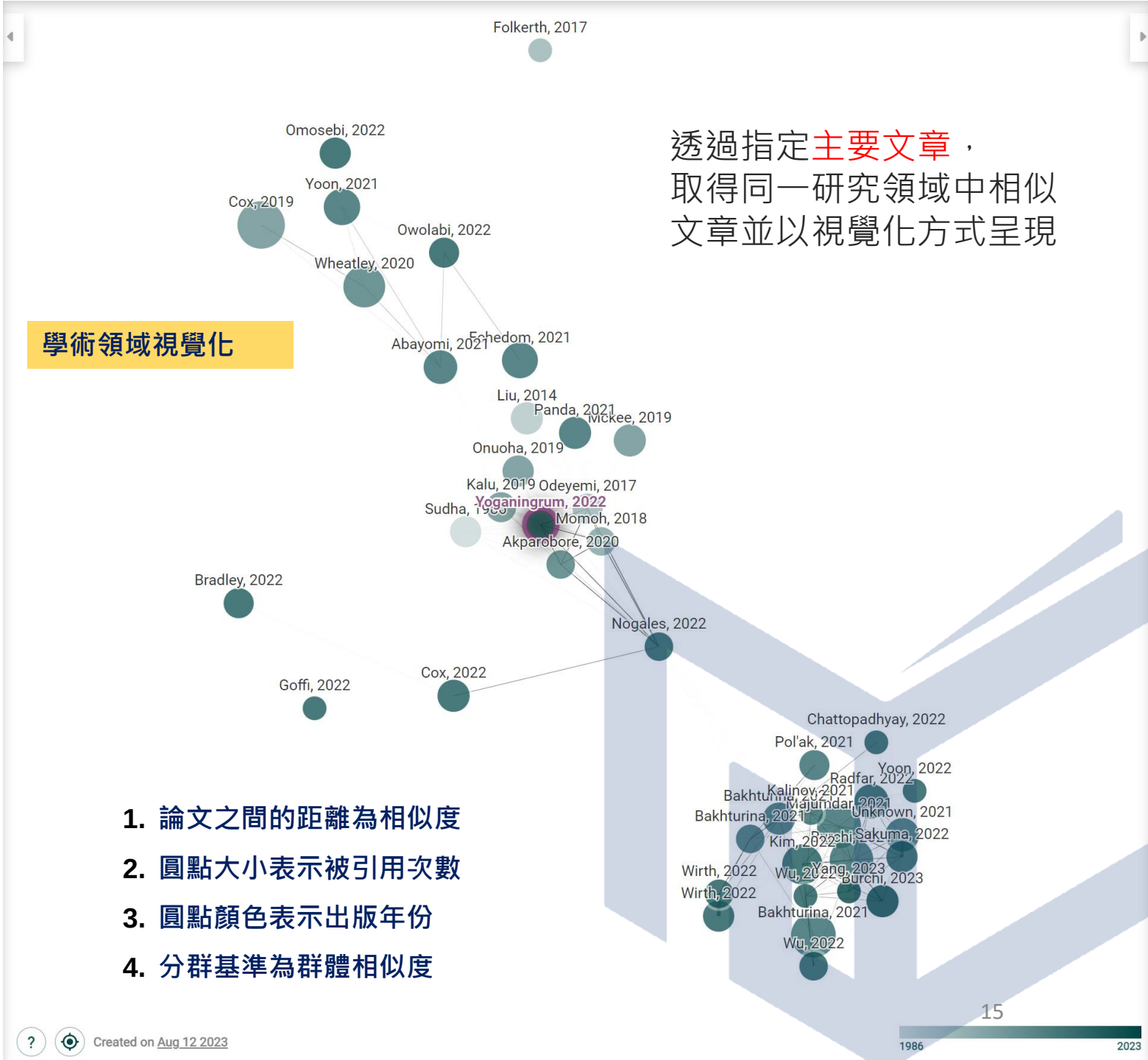
- **User services**
- ✓ Providing **reliable and valuable services** tailored to unique user groups
- ✓ AI tools within their library services
- ✓ Ensure more **personalized** and intuitive services.
- ✓ …

AI應用於圖書館工具

- ✓ Connected Papers
- ✓ Semantic Scholar
- ✓ Hypothesis

Connected Papers

https://www.connectedpapers.com/



Connected Papers

https://www.connectedpapers.com/

- 以原始論文為出發點，根據研究領域、關鍵字、主題、引用關係等基礎因子分析約 5 萬篇文章並選擇與原始論文具有最強關聯性的數十篇文章並製成關係圖
- 透過同被引（Co-citation）與書目耦合（Bibliographic Coupling）計算兩兩論文之間的相似度
- 圖中論文以相似度進行位置排列，即便無引用關係的兩篇文章也互相連結且放得很近
- 透過力導向圖（Force Directed Graph）將相似論文分成一群，並將不太相似的論文推往遠方。與原論文最相近的論文相似度最高，將明顯標記
- 分析資料來源：Semantic Scholar Paper Corpus

Semantic Scholar

分析文章影響力

透過機器學習分析影響引用次數的因素以及文章之間的動態因子

Semantic Scholar

Identifying Meaningful Citations

✓ 透過機器文字處理取得每篇文章中的以下資訊，確認文獻的重要性，進而計算影響

- 直接引用總次數
- 各段落直接引用次數：透過章節定義引用來源
- 非直接引用總次數、各段落非直接引用次數
- 作者重複狀況：文章與文獻出現相同作者時為 True
- 文章認為有用：透過特定片語確認文獻有用時為 True
- 文獻出現在表格或圖表說明中
- 引用文獻數目倒數：引用文獻越少，單一文獻越重要
- 直接引用比例：單一文獻直接引用次數除以所有文獻直接引用次數
- 計算文章與文獻摘要相似度
- 計算文獻 PageRank
- 計算文獻之間產生循環之後，引用同篇文獻的文章數
- 文獻研究領域

Semantic Scholar

自動產生文章摘要

✓ 透過專業背景知識與自然語言技術針對 6000 萬篇文章摘要，以利快速理解內容

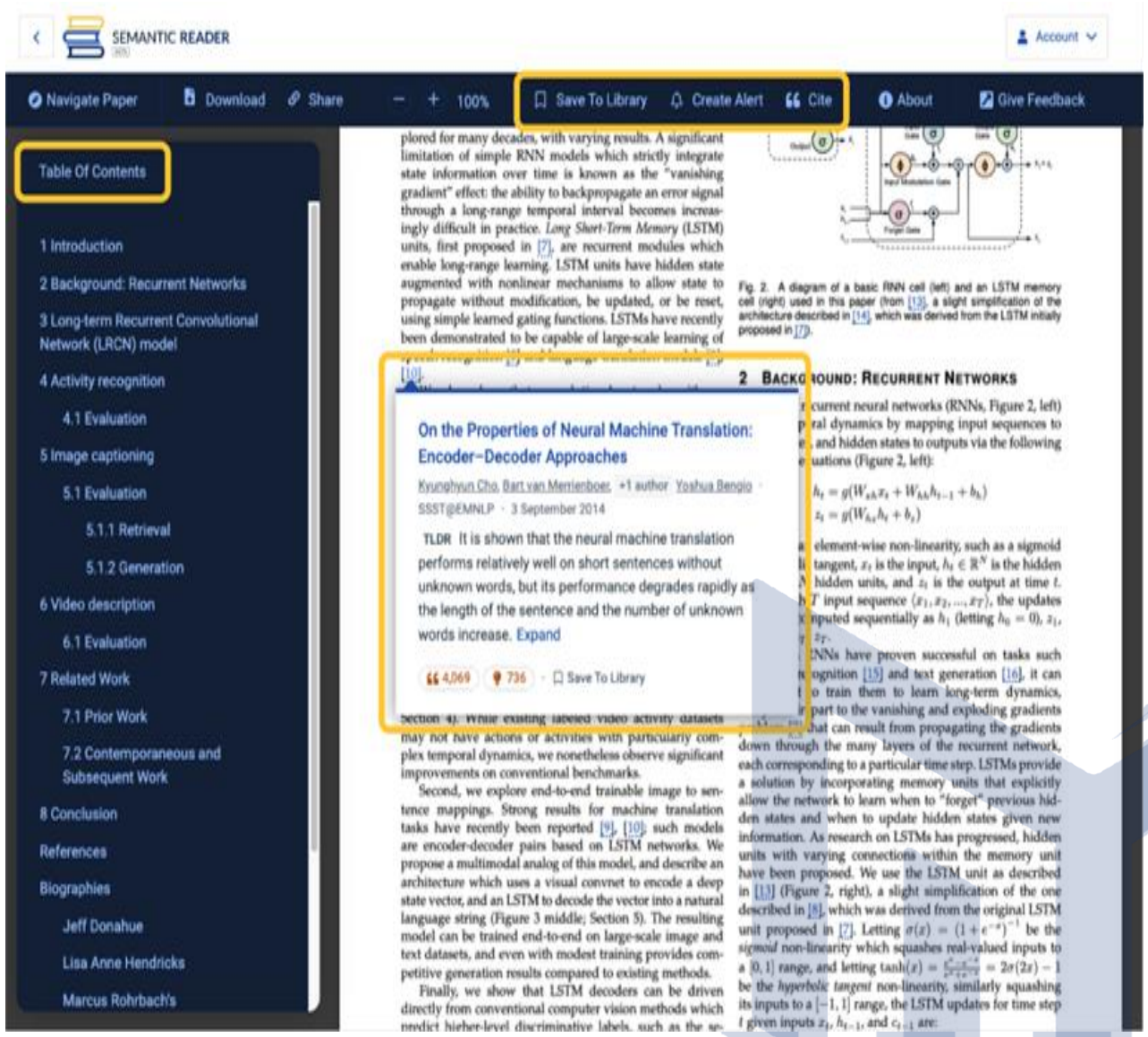
AI 驅動的個人化文章推薦

✓ 透過 AI 學習個人感興趣文章及新進著作間關係，自動推薦最新內容，維持知識更新

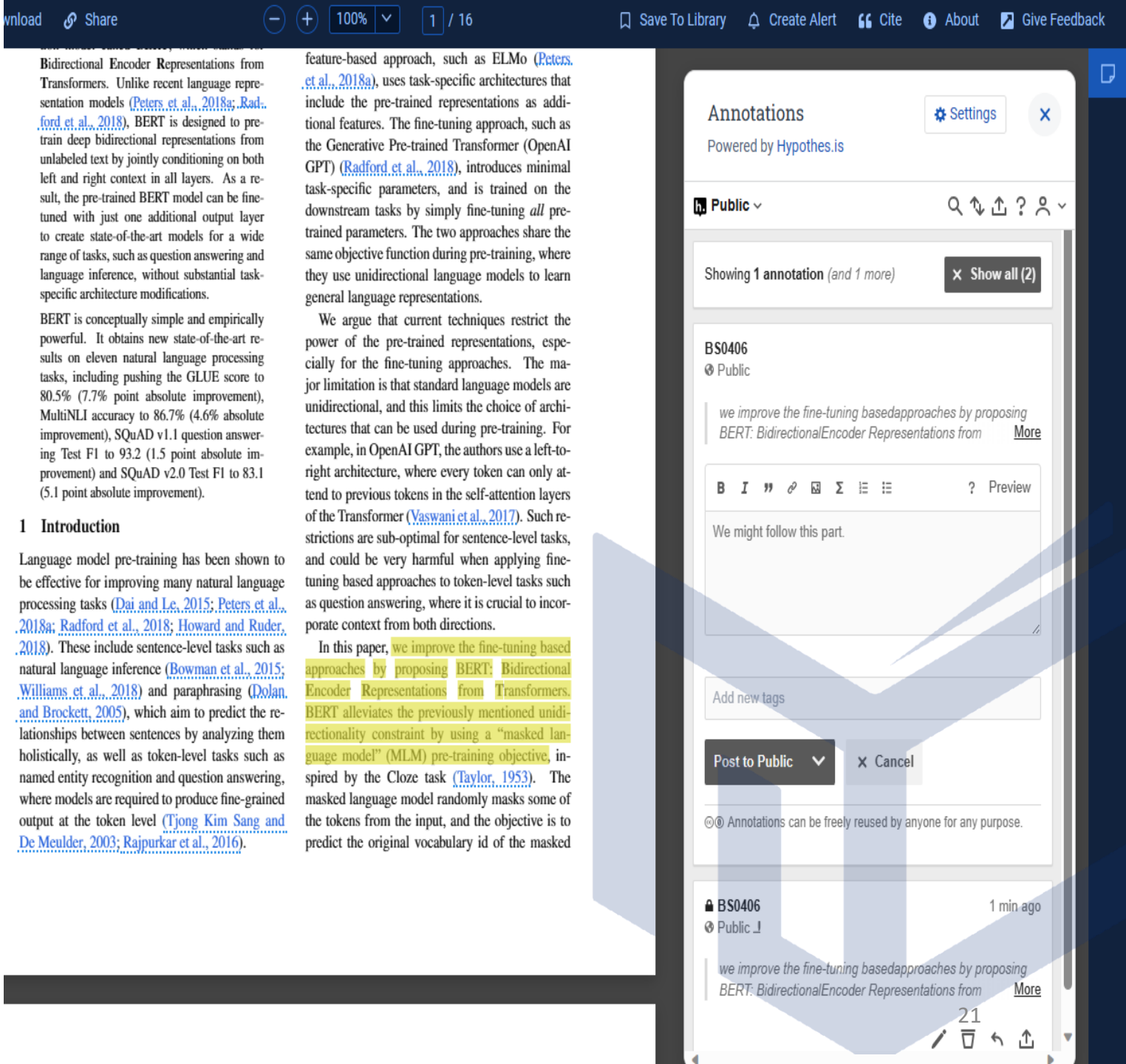
整合式文章閱讀器

✓ 透過分析文章與關聯引用文獻，可直接閱讀引用文獻摘要、查閱文章目錄、放入個人收藏並檢視機器自動標記的重點

Semantic Scholar Reader



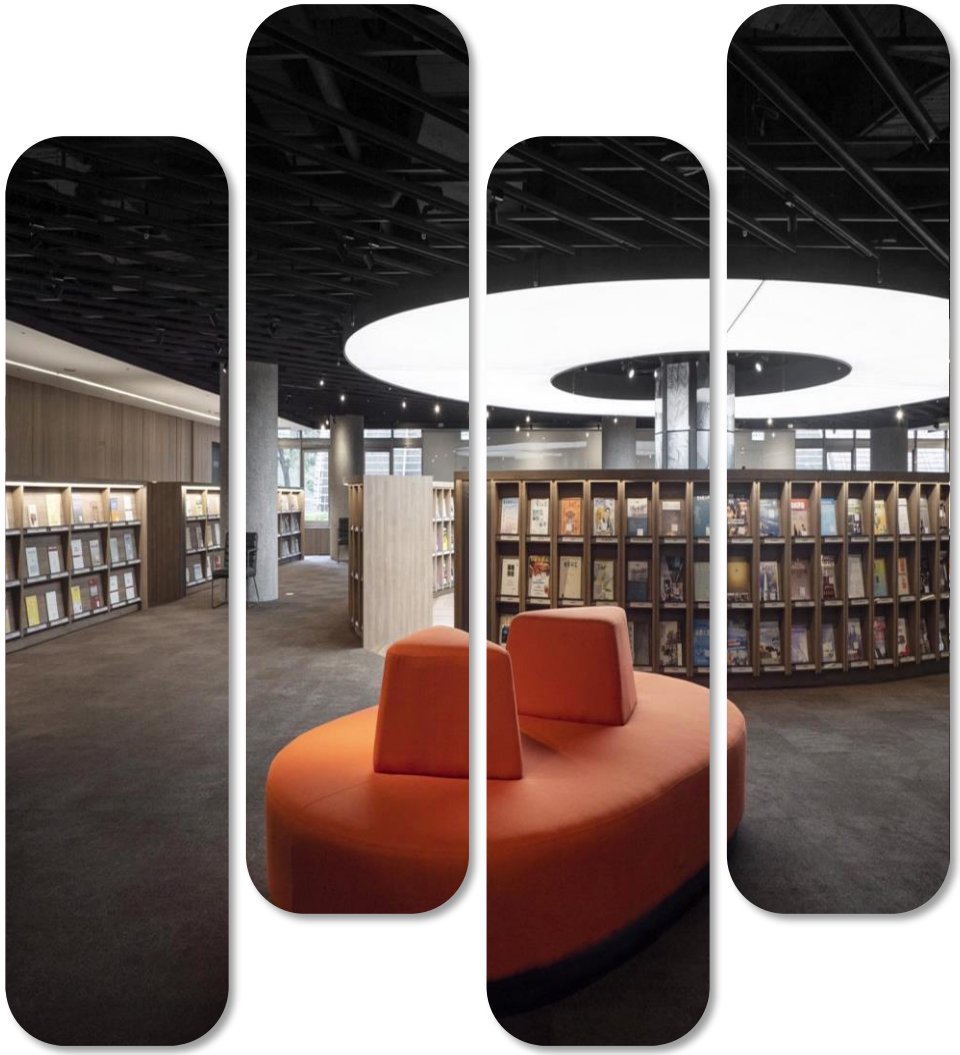
Hypothesis



NYCU Library Value

Aggregation in Library

- 圖書館可檢索近2億筆館藏資源，如何建立平台，引導師生，提供更有系統、更具智慧的「個人化」「精準」資訊服務？
- 提供服務的過程中，同時可保存與再利用這些校園的智慧資產，為主要目的與核心概念



Active Collector and Gate-Opener

For Researchers in Research Life Cycle

讀者的需求

- 面對資訊爆炸的環境，看不完的資料與念不完的文章（資訊焦慮）
- 如何過濾不需要的資料？針對某主題作系統的分析與關聯？Learning Path?
- 是否能(主動)提供可能的研究趨勢？
- 針對主題核心的文章（或圖書）能夠具備「追蹤機制」？（主動提供最新資料）
- …

How Much Information?

校園智慧資產

- ✓ 師生每天在校園中產生無數的有形與無形資料
- ✓ 教學、研究或活動的過程中產生的資料
- ✓ 各式各樣不同的類型資料（文字，照片，影片…）
- ✓ …

保存與再利用

- ✓ 校園智慧資產如何蒐集，組織與保存？如何再利用？
- ✓ 是否有很好的機制，智慧化地主動蒐集？
- ✓ **從資訊生命週期（供應鏈）的概念出發，建置一整合智慧資產平台**

iRDM
@
NYCU

透過平台紀錄校內有形與無形的研究成果與研究歷程

以AI各種工具加值模式協助各階段工作，

以研究生命週期觀點，整合既有資源，

AI, all in one @ Research Life Cycle platform 具體範例

1 研究構思，文獻蒐集

範例工具：Connected Papers、
個人化推薦文章

2 文獻閱讀與筆記

範例工具： Hypothesis、Semantic Scholar Reader

3 研究數據蒐集與管理

既有工具：NYCU Dataverse

4 文章撰寫

範例工具：Mendeley、EndNote、

5 發表策略

範例工具：JCR、WOS、SciVal…

結論與討論

- ✓ 無論時代（科技）如何改變，圖書館與館員的核心價值仍然不變，圖書館館員仍扮演著重要的角色，但須與時俱進
- ✓ 館員是資料正確性的守門員，需要教導讀者資料素養與AI素養
- ✓ 館員更需善用A I的工具，積極改善自己的工作流程、完善規劃與決策品質
- ✓ 利用AI工具，提供讀者更正確與便利的資訊服務，協助讀者更有效率的學習與研究，縮短讀者的學習與研究時間

國立陽明交通大學圖書館

National Yang Ming Chiao Tung University Library

謝謝聆聽

Thank you for listening