月旦知識庫

第41期 2026年9月18日

當20歲的命題大綱遇到AI：用科技開啟新契機

歐陽安琪 考選部測驗發展司專門委員、 范諄佑 考選部測驗研究中心專案研究員

> *“ 考選部透過「AI協助大量整理、專家負責專業判斷」的人機協作，將歷年試題轉化為精進國考的實證資料，為命題大綱持續精進開啟新契機。 ”*

對許多正在準備國家考試（簡稱國考）的人來說，在時間及記憶力有限的情況下，最實際的問題是該如何精準備考。「命題大綱」能替應試科目劃出範圍，讓命題委員有依循的界線，也讓應考人有準備的方向，所以考選部自2005年起，開始有系統地建置公務人員考試各類科的命題大綱，往後陸續擴及數千個應試科目，這套機制運作至今，已經有超過二十年的歷史。

由於國考科目眾多，現今命題大綱修正大多採逐次、局部的方式；其次，歷年國考試題資訊量體龐大，如果要針對成千學科進行知能對應作業，需投入大量學者專家、行政人力、時間及經費等資源，所以全面性的循證分析並不容易。幸運的是，人工智慧（Artificial Intelligence, AI）擅長協助處理大量文字的比對與歸納，使得資料整理及初步分類工作得以加速。本文嘗試讓命題大綱與AI相遇，讓命題大綱系統性檢視具有持續推動的可能。

本文由考試院授權提供

月旦知識庫

命題大綱建置工程：一段耗時費工的歷程

早在1999年，考選部便率先從專技醫事人員考試之參考用書著手，於2001年逐步發展為醫師考試的命題大綱。而大規模有系統地命題大綱建置工程，始於2003年4月，考試院決議請考選部就各項考試新增之應試專業科目，先行研擬命題範圍，供典試委員會參考，這項工作也在2005年起正式成為考選部的重要業務之一。

命題大綱建置過程，係基於各類科職務的專業知識及核心能力，綜整學術及實務界的意見，鉅細靡遺地列出該考科學科領域的命題範圍，並且可以命擬相關的綜合性試題。然而，國考涵蓋的類科範圍龐大，當時公務人員考試和專技人員考試合計超過700個類科、3,000個應試科目。在人力與經費有限的情況下，考選部採取分年、分類科的方式辦理。每次命題大綱的制定，須經資料蒐集及分析、座談與審查，往往需時3至6個月，甚或更久才能取得共識，得以定案公告。

命題大綱的定位：應考人與命題委員的定海神針

命題大綱之所以耗時耗力仍然要做，是因為它同時承擔著兩個重要功能。

對應考人而言，它是備考時的方向指引。面對應試科目浩瀚的專業領域，如果沒有明確範圍，準備考試容易流於漫無邊際，徒增時間與心力負擔，增加備考成本。命題大綱把考試範圍具體例示出來，讓應考人把力氣用在對的地方。

對命題委員而言，它是維護考試信度與效度的依循標準。在建立命題大綱之前，國考偶因各年度命題委員的學術背景不同，導致試題難易度出現落差。命題大綱依據該職務應具備的核心知能與職系說明書擬訂，為命題委員提供了公正、客觀且具備法制基礎的依循標準，也讓不同年度間的試題水準得以維持一致。

二十年之後：知能翻新帶來的現實挑戰

理想的指引必須是「與時俱進」，此指引如採用符合當前趨勢的架構，將能協助應考人把時間精準投注在最具核心價值與效益的內容上。但若依循過往

本文由考試院授權提供

月旦知識庫

「全面委託專家研究、召開產官學座談與期末審查」的模式，面對數千個科目的調整，所需的時程、經費及人力需求，考選部恐難以負荷。而目前考選部的作法，除了配合科目整併而訂定新的命題大綱外，主要仰賴各項考試典試委員會、試題疑義會議，或題庫委員在參考運用過程中發現不合時宜之處，個別的提出局部修正建議，再由考試業務單位微幅調整；至於整份命題大綱是否適切對應各學科領域應具備的核心知能，亦或對應考人真正達到備考引領的功能，則較少有系統性的檢視。現在正是時候，透過AI新科技一探究竟。

迎向未來：AI「人機協作」重塑動態指引新契機

面對知能翻新與行政資源受限的雙重夾擊，考選部嘗試借重AI，推動「人機協作」。須先說明的是，本文作為一個議題的初探，只是把這次實作的歷程、方法與數據攤開來，提供關心考政的讀者一個思考空間。再者，AI在這裡扮演的不是取代專業判斷的角色，而是一個有效率的數位助理，讓AI先做粗重的比對工作，把專家的時間留給真正需要判斷的地方。

導入AI的過程並非一次到位。初期測試時，系統的表現並不穩定，有時AI被字面相似牽著走，例如把「感染性肺炎」的題目歸到呼吸系統，而非感染疾病；有時又會過度自信，生成不存在的分類編號。團隊因此必須透過提示詞，反覆與AI溝通調整，說明工作紀律：「先看整體脈絡再看細節、答案只能從候選清單中選、不確定時要老實說不確定」。經過幾輪測試與修正，AI的檢索與分類才逐漸穩定下來。

讀者也許好奇，這套框架實際上是怎麼搭起來的？考選部的團隊在Claude（Anthropic公司的AI工具）中建置一個專案，專案裡先放一份「專案說明檔」，用白話文告訴AI這個專案的目標、資料放在哪裡，以及前述那幾條紀律；接著把整個工作拆解成四個「技能模組」（Skill：為考綱節點撰寫定義、檢索候選節點、兩步驟分類推理、統計與產出報表），每個技能模組就是一份寫給AI的標準作業程序，明確規定輸入、步驟與輸出格式。最關鍵的是「系統提示詞」的設計，裡面只寫結構規則，例如「定義必須能與同層節點區隔」、「答案只能從候選清單中選」，完全不寫任何測試科目的專業知識。各科目的通用性，正是這樣達成的。這樣的設計，也意味著本框架不限

本文由考試院授權提供

月旦知識庫

於單一案例，只要備妥命題大綱檔案，即可快速套用至不同科目，提高跨類科、跨科目應用的可行性。

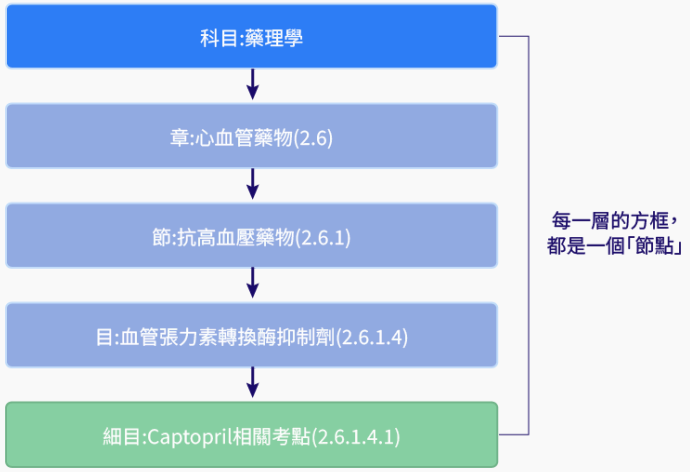
考量國考試題題型分為申論題和測驗題，題數不同，關鍵知識點出現的頻率不同，而其所評測的專業能力面向亦不相同，所以本次特別從公務人員考試及專技人員考試，分別挑選申論題及測驗題的科目，依其特性做為初探的出發點。

案例一：測驗題指引的收斂，從茫茫題海中找方向

第一個試驗場，選擇了護理師考試的「基礎醫學」科目。之所以從這裡切入，是因為在導入AI之前，考選部透過「護理師國考題務精進計畫」，已經先請專家針對這個科目，把原本只有兩層的命題大綱，逐層討論、延伸出第一、二、三層次的次條目，並從歷屆考畢試題中挑選出的610道國考核心職能例題，逐一標註到正確的次條目上。換句話說，這裡已經有一份現成的「專家答案」可以核對，此份資料正可以做為檢驗AI準確度的標準答案。

這份經專家細緻化後的知能分類架構，總共展開成近一千四百個「節點」。所謂節點，指的是這棵命題大綱樹狀圖上，每一個分岔處與末端的位置，好比一本書的目錄，「章」底下有「節」、「節」底下有「目」，每一條目錄項目就是一個節點，越往末端，知識點越具體、越貼近實際考點（如圖1）。AI的任務，就是把每道只有短短數行的題目，沿著這棵樹一路走到正確的節點位置。

圖1 命題大綱的節點階層示意
(以護理師考試「基礎醫學」科目為例)

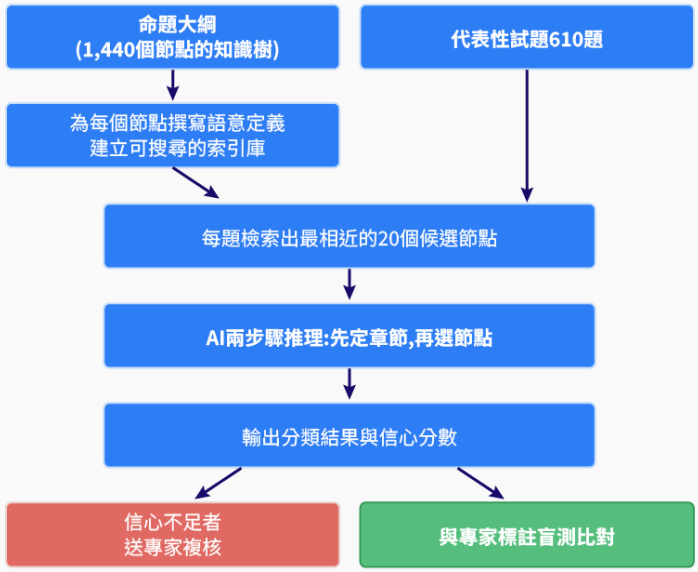


資料來源：作者自行整理。

本文由考試院授權提供

有了專家已完成標註的610題可作為標準答案，團隊接著設計一套可重複使用的分類流程（如圖2）：先為每個節點撰寫一段簡短的「語意定義」，建成可用語意搜尋的索引庫；每道題目進來，先檢索出最相近的候選節點，再交給AI推理定位，並逐層附上信心分數。

圖2 測驗題考綱分類對應流程



資料來源：作者自行整理。

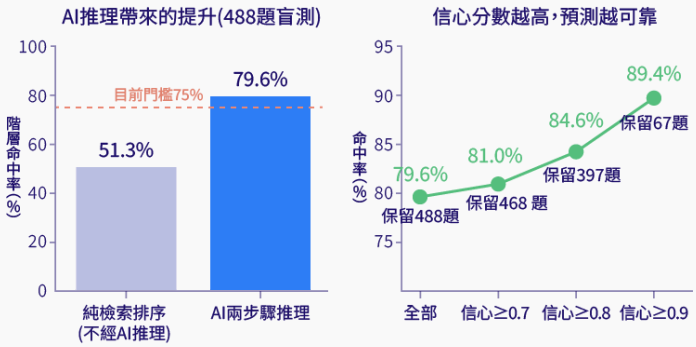
AI的表現如何？在測試設計上，團隊採用「二八分」，將610題分成兩部分：二成（122題）作為開發集，用來校準提示詞的寫法與檢索參數；八成（488題）作為盲測集，供最終驗證，且事先不讓AI接觸這部分的專家標註答案。這裡的開發集，並不是要讓AI「學會」這個科目的專業知識；AI本身已具備廣泛的通用常識與語言理解能力，開發集只是用來確認分類流程的設計是否合理、檢索出的候選節點能不能抓到重點，而不是額外灌輸某個科目的專業知識。這個案例整體命中率達79.6%，超過團隊事前設定作為初步可用性評估的75%目標。一般若是不經過AI推理，單純依檢索分數排序，命中率僅51.3%。此一結果顯示，讓AI「理解題意再判斷」而非「照搜尋排名抄答案」，是提升準確度的關鍵（圖3左）。信心分數同樣具有參考價值，門檻拉得越高，保留下來的預測命中率也越高（圖3右）。以信心0.7為界，488題中僅有37題低於此門檻，占比7.6%，此部分即建議送交專家複核。

案例一的意義在於驗證AI可協助完成大部分試題的初步分類工作，在準確性可接受的情形下，專家可優先將心力集中於低信心案件，並對其他分類結果

本文由考試院授權提供

抽核，形成「AI協助大量整理、專家負責專業判斷」的人機協作模式，大幅降低盤點作業的人力成本。

圖3 測驗題盲測成果:AI推理的貢獻與信心分數的可靠性



資料來源：作者自行整理。

案例二：申論題指引的活化，描繪出題版圖

在申論題部分，本研究選擇公務人員高等考試三級考試都市計畫技術類科「都市及國土計畫理論與法制」科目作為分析案例。該科目採申論式題型，每次考試共4題，每題配分25分，具有申論題題數有限、重視綜合分析與論述能力之典型特性。另一方面，其命題大綱涵蓋都市計畫、國土計畫及相關法制等內容，並採分層架構設計，便於將歷年試題逐題對應至命題大綱細目，分析各細目之實際命題分布情形，提供申論題命題大綱檢視之實證參考。

本次分析資料取自考選部全球資訊網資料開放專區之「國考、等級、類科組別、節次、科目名稱、試題型態及考畢試題（含測驗題標準答案）連結網址開放資料」，其連結之最早年度為2012年科目試題，因此分析範圍定為2012年至2025年，其中包含公務人員高考三級考試、地方特考三等考試、鐵路特考三等考試都市計畫技術類科相關科目共231道申論題。需說明的是，2012至2023年，「都市及區域計劃法令與制度」（簡稱法令科目）與「都市及區域計劃理論」（簡稱理論科目）兩科分別列考，各自有一套命題大綱，這段期間累積法令科目107題、理論科目108題；2024年起兩科合併為「都市及國土計畫理論與法制」（簡稱理論與法制科目），適用新訂的命題大綱，至2025年累計16題。三套大綱彼此獨立分析、互不混用（各年度題數與對應大

本文由考試院授權提供

月旦知識庫

綱如圖4）。想回答的問題是：這231道申論題究竟觸及了命題大綱中哪些節點？哪些節點從未被考過？

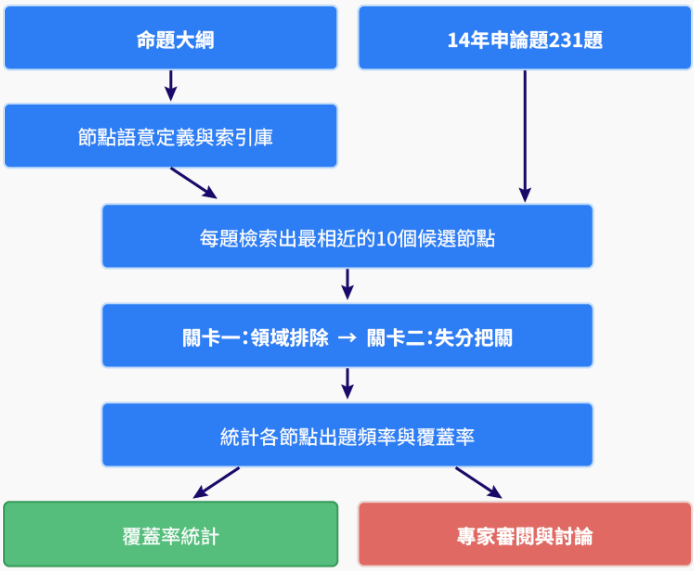
圖4 都市計畫公務人員考試技術類科(都市及國土計畫理論與法制)科目沿革



資料來源：作者自行整理。

由於申論題沒有像案例一「基礎醫學」科目已有專家標註審定的知能架構可供參考檢驗，這部分的分析更像是初探中的初探，目的在於先建立一套可行的判讀方法，而非驗證準確率。判讀設計為兩道關卡（如圖5）：第一道先依試題的核心領域，排除掉語意相似但實際不相關的候選節點；第二道則用「應考人若不熟悉這個節點（知識點），會不會嚴重影響此科目的得分？」這個具體的問題把關。為了把「這道題目真正在測驗的知識點」和「答題時可能用到、但非重點的背景知識」區分開來，避免只要語意相近就照單全收，讓標記結果更貼近試題實際考核的內容。符合這個判準才標記，只是背景知識則不標記。

圖5 申論題考綱覆蓋率分析流程



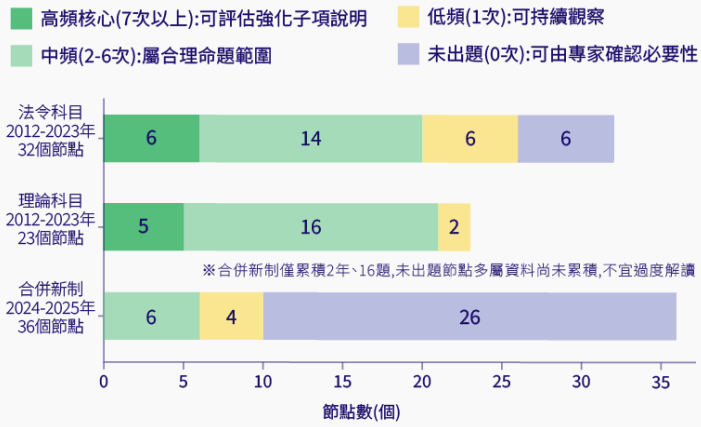
資料來源：作者自行整理。

本文由考試院授權提供

公職知識庫

案例二的分析結果整理如圖6。團隊將各節點依12年（合併新制為2年）的出題頻率分為4層，直接對應後續檢視的優先順序。以法令科目而言，32個節點中有6個屬高頻核心（7次以上），其中「都市計畫法」單一節點就出現31次，平均每三題就有一題與它相關，此類節點可評估是否提供更細緻的子項說明，幫助應考人精準備考；中頻14個屬合理命題範圍；至於6個在這12年間從未出題的節點，則可由專家確認其於當前執業的必要性。理論科目的分布更為集中，23個節點全數曾出題，無零頻節點。2024年起兩科合併的理論與法制科目，因僅累積2年、16題的資料，尚有26個節點未出題，屬資料尚未累積的自然現象，不宜過度解讀。

圖6 出題頻率分層:命題大綱檢視的優先順序參考



資料來源:作者自行整理。

結論：建立可行框架，逐步擴大應用

回顧這兩個案例，AI確實加快了整理與比對的速度，過去需要數月才能完成的人工盤點，如今在幾天內就能產生初稿，惟每一個節點的定義是否準確、每一筆低信心的分類該怎麼判斷，仍然仰賴專家的審閱與確認。AI縮短的是前期作業的時間，取代不了的是專業判斷這一關。下一步，希望能將這套框架逐步延伸到其他類科，並持續累積實際運作的經驗，讓命題大綱的檢視工作，能用更有效率的方式持續進行下去。期待透過人機協作的核心知能架構，未來可做為命題大綱補充資料文件，或是提供命題大綱精進的實證依據，以減少應考人備考的壓力及成本。

延伸閱讀

本文由考試院授權提供

月旦知識庫

· 國家考試電腦化測驗使用多媒體科技命題之研究 / 曾憲雄、劉湘川、黃國禎、 郭伯臣、翁瑞鋒 / 國家菁英季刊第5卷第2期

上一篇	回前頁	下一篇

本文由考試院授權提供