

現代科學探索

目 錄

- 1 海南熱帶水果冷鏈運輸中溫度波動對果實品質與貨損率的影響
黃 楠
- 12 AIGC生成內容的權威性與品質控制科學知識生成中的驗證與評價機制探討
陳 明
- 30 分層土壤中降雨驅動下重金屬遷移行為的層間擴散機制探討
趙天宇
- 37 推薦系統中演算法強化與使用者偏好共演化機制研究
周 波
- 49 國產AI晶片上的圖優化編譯技術研究——以昇騰CANN為例
周映竹

DOI: 10.53104/xdkx-ts.2025.01.02.001

海南熱帶水果冷鏈運輸中溫度波動對果實品質與貨損率的影響

黃楠¹

1. 海南大學，海南 海口 570228

摘要：海南是我國重要的熱帶水果種植區域，核心品類包含芒果、荔枝、火龍果、菠蘿蜜和蓮霧等，相關產品暢銷國內外市場。冷鏈物流是保障水果長途流通品質的核心環節，溫度不穩定仍是突出難題，其會加快果實呼吸代謝，引發冷害或熱害，助長微生物繁殖，縮減貨架期限，提升損耗比例，造成明顯經濟損失。本文剖析海南熱帶水果冷鏈物流中溫度不穩定對水果品質及貨損比例的影響機制，探究不同溫度波動程度對應的理論體系及損耗路徑，參考發達國家與熱帶水果出口國的溫度管控實踐經驗。基於上述分析，本文提出針對性溫度管控優化方案，強化溫度穩定性，減少損耗量，提升海南熱帶水果供應鏈的綜合競爭力。

關鍵字：冷鏈運輸；溫度波動；熱帶水果；品質保持；損耗率；海南

The Impact of Temperature Fluctuations on Fruit Quality and Loss Rate in Cold Chain Transportation of Tropical Fruits in Hainan

HUANG Nan¹

1. Hainan University, Haikou 570228, P.R.China

Correspondence to: HUANG Nan; Email: huangnannn@yeah.net

Abstract: Hainan is an important tropical fruit-growing region in China, with core varieties including mango, lychee, dragon fruit, jackfruit, and wax apple. These products are popular in both domestic and international markets. Cold chain logistics are crucial for ensuring the quality of fruits during long-distance transportation, yet temperature instability remains a major challenge. Such fluctuations can accelerate fruit respiration and metabolism, cause chilling or heat damage, promote microbial growth, shorten shelf life, increase loss rates, and result in significant economic losses. This paper analyzes the mechanisms through which temperature instability in Hainan's tropical fruit cold chain logistics affects fruit quality and loss ratio, exploring the theoretical framework and loss pathways associated with different degrees of temperature fluctuation, while drawing on temperature control practices from developed countries and major tropical fruit exporters. Based on this analysis, the paper proposes targeted temperature management optimization strategies to enhance temperature stability, reduce losses, and improve the overall competitiveness of Hainan's tropical fruit supply chain.

Key words: cold chain transportation; temperature fluctuation; tropical fruits; quality preservation; loss rate; Hainan

收稿日期：2025-08-14 返修日期：2025-12-04 錄用日期：2025-12-12 出版日期：2025-12-22

通信作者：huangnannn@yeah.net

引用格式：黃楠. 海南熱帶水果冷鏈運輸中溫度波動對果實品質與貨損率的影響[J]. 現代科學探索, 2025, 1(2): 1-11.

引言

海南是我國重要的熱帶水果種植區域，重點出產芒果、荔枝、火龍果、菠蘿蜜、蓮霧等優質特色品類，近年海南自貿港建設持續推進，跨區域農產品流通效率不斷提升，熱帶水果除滿足本地消費需求外，還大量銷往內陸地區並出口至東南亞、中東等海外市場，對運輸時效與品質保障提出更為嚴格的標準。

冷鏈物流是維繫熱帶水果長距離流通品質的核心環節，相關研究顯示，健全的冷鏈體系能將果蔬腐損率從常溫運輸狀態下的 15% - 30% 控制在 5% - 10% 的較低水準，海南熱帶水果冷鏈運輸實際運營中仍存在明顯短板，冷藏車輛保有數量不足、預冷設施佈局不均衡、運輸規範缺乏統一性等問題表現突出，溫度波動作為影響水果品質與貨損比例的關鍵不穩定因素，在裝卸作業、中途停靠、製冷設備故障等環節極易出現暫態升降現象。

溫度的不穩定變化直接加快果實呼吸作用與代謝進程，還可能誘發冷害問題、促進微生物滋生，最終導致果實口感變差、色澤退變、腐爛概率上升，以芒果為例，運輸環境溫度在 10℃ 左右時，波動幅度超出 $\pm 3^{\circ}\text{C}$ ，果皮就容易出現褐斑且軟化速度明顯加快，對於荔枝這類高呼吸強度的水果，溫度短時間內升高會大幅加速褐變反應，使其商業可售週期縮短 1 - 3 天，這些品質層面的改變影響消費者食用體驗，還直接加重供應鏈各環節的經濟損耗。

當前國內外關於溫度波動的研究多集中在儲藏環節，針對冷鏈運輸過程動態溫度管控的系統性研究相對匱乏，海南產業實際中，水果運輸距離通常較長，海口至北京的運輸里程約 3000 公里，運輸週期需要 3 - 4 天，且運輸過程涉及輪渡、鐵路、公路等多種方式銜接，使得溫度波動風險更具複雜性與多樣性，深入探究溫度波動對熱帶水果品質與貨損率的影響機理，結合海南冷鏈運輸的獨特特點提出針對性溫控優化建議，對增強海南水果供應鏈競爭實力、減少經濟損失具有重要的理論價值與實踐意義。

本文以海南主要熱帶水果為研究物件，分

析冷鏈運輸環節溫度波動對果實品質及貨損率的影響路徑，借鑒國內外冷鏈管理實踐經驗，提出適配海南產業特徵的溫控管理優化策略，為海南熱帶水果冷鏈體系的完善提供參考依據。

1 海南熱帶水果冷鏈運輸現狀與問題分析

海南地處熱帶季風氣候帶，全年高溫高濕，適宜各類熱帶水果培育，近年依託海南自貿港相關政策支持及交通基礎設施持續完善，熱帶水果產業快速發展，海南省統計局資料顯示，2023 年全省熱帶水果總產量突破 470 萬噸，芒果、荔枝、火龍果的產出規模位居全國前列^[1]，消費需求升級與跨區域電商銷售崛起推動海南水果大量湧入珠三角、長三角、京津冀等遠距離消費市場，對運輸品質保障的依賴度明顯提升。

冷鏈物流是維繫熱帶水果長途運輸品質與市場價值的核心，相較於常溫運輸，冷鏈能顯著延緩果實呼吸強度與成熟進程，減少果實腐爛與水分流失，從全產業鏈視角，海南熱帶水果冷鏈運輸水準仍存在顯著短板。

冷鏈基礎設施佈局失衡

海南冷藏車輛保有量偏少，行業協會統計資料顯示，2023 年全省冷藏車總量不足 2500 輛，且多集中於海口、三亞等核心城市，部分縣域缺乏專業冷鏈運輸支撐能力，產地預冷設施普及比例較低，不少水果採收後長時間處於常溫環境，未能完成“採收後即時預冷”的標準化處理流程^[2]。

運輸環節溫度波動現象頻繁

運輸線路特性導致海南水果冷鏈常需跨海輪渡或鐵路中轉，港口或網站裝卸作業時，運輸車輛需長時間關閉製冷裝置，車廂內部溫度短時間快速上升，夏季高溫時段溫度上升速率可達每分鐘 $0.3 - 0.5^{\circ}\text{C}$ ^[3]，部分老舊冷藏車輛製冷性能不足，溫度控制精度較差，難以維持穩定的低溫運輸環境。

溫度管控標準尚未統一

當前海南熱帶水果冷鏈運輸環節未建立統一行業溫控規範，不同企業多憑自身經驗確定運輸溫度，例如芒果運輸溫度，部分企業設定為 12℃，另有企業設定在 10℃ 或更低區間，缺乏依託果實生理特徵的科學支撐，這種溫度設定差異影響運輸品質，增加行業整體管理難度。

經濟成本與管理水準制約溫控效果

部分中小型運輸企業為壓縮運營開支，縮短製冷裝置運行時長或混裝不同品類水果，導致車廂內溫度分佈不均，司機與裝卸作業人員溫度管控意識薄弱，對溫度波動危害缺乏足夠認知，使得運輸全過程溫控執行存在諸多漏洞。

總體而言，海南熱帶水果冷鏈運輸近年取得一定進步，溫度波動問題仍普遍存在，成為影響水果品質與貨損比例的關鍵風險要素，後續研究需深入剖析溫度波動的作用機理與損耗路徑，結合海南地理特徵與物流行業特點，探尋科學高效的溫控優化路徑。

2 溫度波動影響果實品質的機理

海南熱帶水果採收後仍維持複雜生理代謝過程，涵蓋呼吸代謝、乙烯生成、細胞分裂及組織修復等，這些過程的活躍程度與環境溫度密切相關，冷鏈運輸的核心工作之一是通過穩定適配的低溫條件，抑制不必要的代謝反應，延緩果實成熟與衰老進程，實際運輸場景中，溫控設備性能起伏、裝卸作業環境暴露、能源供應波動、人為操作偏差等因素，常造成運輸環境溫度變動，對熱帶水果品質形成多維度影響。

環境溫度升高，果實呼吸強度與乙烯釋放量呈指數級上升，引發一系列連鎖效應，細胞內有機酸快速消耗，糖分代謝進程加快，風味成分平衡狀態遭到破壞，果肉出現酸度降低、甜味過度積聚、芳香物質散失，同時酶促反應速率提升，促使果肉軟化、褐變等衰敗現象提前發生，芒果、荔枝、龍眼這類高糖高水分熱帶水果，短時間溫度攀升就足以明顯縮短商品流通週期^[4]。

溫度降至臨界值以下，部分熱帶水果顯現冷害症狀，香蕉在 12℃ 以下易出現果皮暗褐、果肉失水，番木瓜處於 7℃ 以下環境會產生水

浸狀斑點及組織壞死，冷害伴隨細胞膜脂質相變，膜結構喪失流動性與完整性，細胞通透性增加，胞內酚類物質與多酚氧化酶接觸，引發快速褐變，這些損傷在回溫過程中往往進一步加重，細胞已失去正常滲透調節能力，呼吸與蒸騰速率異常升高。

溫度頻繁變動進一步加劇果實損傷，反復冷熱交替造成細胞膜多次膨脹與收縮，加速膜蛋白降解及脂質過氧化，導致果實質構劣化、組織鬆散，這種環境變化還誘發次生代謝產物異常累積，如苦味物質、代謝毒素等，影響消費者感官接受程度。

微生物生態層面，溫度波動極易打破冷鏈運輸中微生物群落相對平衡，溫度處於病原微生物適宜生長區間（15 - 25℃）時，炭疽病菌、青黴、灰黴等腐敗真菌及部分細菌快速增殖，即便後續降溫至抑制範圍，這些微生物仍可能以休眠孢子形式存活，在終端銷售或消費者儲藏階段重新活躍，顯著提高果實腐爛風險與貨損比例。

海南熱帶水果受溫度波動影響更為複雜嚴重，海南出口類水果如芒果、火龍果、菠蘿蜜等，往往需要經歷跨島運輸與長距離物流，涉及港口集散、船舶運輸、冷庫中轉及多次裝卸作業，這些環節中溫控中斷的累積效應，常使運輸全程平均溫度波動幅度達到 2 - 5℃，這一區間已足以明顯縮短果實貨架週期，海南濕熱氣候條件下，冷鏈一旦出現管控缺口，果實溫升速度極快，微生物繁殖適宜視窗隨之擴大，造成的品質下降基本難以逆轉。

溫度波動是影響海南熱帶水果運輸品質的技術性難題，更是決定其國內外市場競爭能力的關鍵風險要素，維持溫度穩定並降低波動幅度，已成為海南熱帶水果冷鏈體系中提升品質保持率與減少貨損比例的核心任務。

3 溫度波動與貨損率關係的理論分析

3.1 溫度波動與貨損率的相關性框架

海南熱帶水果冷鏈運輸中，溫度變動與貨損比

例存在極強關聯性，這種關聯既表現為直接物理損害，還會通過多重生理生化反應間接影響果實品質，熱帶水果多為高水分高糖分且組織結構嬌嫩的農產品，適宜儲運的溫度範圍較窄，對溫度變動的敏感程度遠超溫帶水果，運輸期間溫度出現異常變動，果實呼吸強度會快速提高，能量消耗加速，可溶性固形物含量降低，酸度失衡，推動果實衰老與軟化^[5]。

溫度變動對貨損比例的作用呈現階段性與累積效應，運輸初期，短時間小幅溫度變動或許僅造成果實內部代謝輕微加快，外觀無明顯改變，這類早期微損傷會在後續運輸及銷售環節逐步放大，使保鮮週期顯著縮短，運輸中後期，多次溫度變動疊加的情況下，果實組織細胞膜通透性降低，抗病能力減弱，極易遭受真菌等微生物侵染，形成明顯腐爛區域，這一過程往往呈現非線性特徵，貨損比例在溫度變動累積到特定程度後會急劇上升。

冷鏈系統實際運行中，溫度變動可能源於多個環節，裝卸時車廂或冷庫門頻繁啟閉、冷機故障或能耗調節不當、運輸途中環境溫度驟變、車載溫控設備監測精度不足等，這些因素常在不同階段相互疊加，使溫度變動成為影響貨損比例的系統性風險因素。

構建關聯性框架需引入多維參數展開分析，溫度變動的幅度、持續時長、發生頻次，果實適宜儲運溫度範圍及耐受臨界值，不同品種的呼吸強度與耐冷敏感程度等，通過這些因素的組合可建立預測模型，將溫度變動轉化為貨損比例的量化指標，鳳梨、芒果等呼吸躍變型水果，溫度每升高 2℃ 可能使貨損比例提升 5%—8%，椰子等耐儲性較強的水果，受影響幅度相對較小。

該框架的應用價值在於，能闡釋溫度變動與貨損比例之間的因果機制，還能夠為運輸環節的風險評估與幹預提供量化依據，通過持續監測與歷史資料分析，可提前識別高風險運輸條件，採取優化裝卸作業流程、提高溫控設備回應速度、設置動態溫度預警等預防措施，最大限度降低溫度變動引發的經濟損失與品質下滑。

3.2 不同溫度波動幅度的損耗路徑

海南熱帶水果對冷鏈運輸中的溫度穩定性

依賴度極高，不同幅度的溫度變動會經由不同生理生化路徑作用於果實，形成多樣化損耗路徑，這些路徑不僅取決於溫度變動的幅度，還與波動持續時長、發生頻次及水果品種特性密切相關^[6]。

輕度波動情境下，溫度變動幅度一般在 $\pm 1^{\circ}\text{C}$ 範圍內，持續時長較短，對果實的影響主要表現為呼吸速率與水分蒸發速率的微弱提升，此時損耗多體現為品質緩慢下滑，例如果皮光澤減弱、風味物質輕微散失，外觀仍能維持較好商品性，這類輕度損傷在銷售末期可能轉化為明顯軟化與失水，加速果實喪失市場競爭力。

中度波動情境下，溫度變動幅度介於 $\pm 2^{\circ}\text{C}$ 至 $\pm 4^{\circ}\text{C}$ 之間，果實代謝活動顯著加快，呼吸消耗大量有機物質，組織細胞膜穩定性遭到破壞，芒果、火龍果等呼吸躍變型水果，這一溫度變動區間可能引發提前成熟，導致硬度下降、糖酸比例失衡，同時微生物繁殖速度明顯提高，果實更易在運輸中或銷售初期出現病斑與腐爛現象，大幅提升貨損比例。

重度波動情境下，溫度變動幅度超出 $\pm 5^{\circ}\text{C}$ ，特別是從低溫向高溫的急劇轉變，會導致果實發生不可逆轉的品質劣變，香蕉、木瓜等部分熱帶水果，劇烈溫度變動可能引發“冷害”或“熱害”反應，表現為果皮褐變、斑點加速擴散、果肉水漬化等，這類損傷通常發生在波動後的 24—48 小時內，幾乎無法通過後續冷卻或保鮮手段逆轉，造成的貨損不僅是數量減少，更包括商品價值的大幅貶損。

不同溫度波動幅度對應的損耗路徑還可能形成疊加效應，運輸初期出現的輕度波動若未得到控制，中後期疊加一次中度或重度波動，會使早期潛在損傷快速放大，導致品質在短時間內集中下滑，這種“鏈式反應”是冷鏈運輸風險管理的重點關注對象，需在監測與幹預中特別防範。

這種分級化的損耗路徑分析，有助於在冷鏈運輸中制定差異化風險控制策略，針對輕度波動，可通過縮短流通時長、優化末端配送速度減少影響，針對中度與重度波動，則需強化源頭裝車溫控、提升車輛隔熱性能，確保溫控設備全程維持高精度運行，避免因一次性失控

造成大規模品質與經濟損失。

3.3 成本—損耗平衡理論在冷鏈管理中的應用

海南熱帶水果冷鏈運輸中，溫度控制的穩定性直接影響運輸環節損耗程度，溫度管控品質與運輸投入成本緊密關聯，成本—損耗平衡理論核心是運輸環節需在保鮮投入與品質損失之間探尋最佳均衡狀態，讓綜合成本（運輸投入成本 + 損耗成本）降至最低，實現經濟收益與品質保障的最優匹配。

這一理論框架下，運輸全過程的綜合成本包含兩方面內容：

其一為管控成本，涵蓋冷鏈裝備採購及日常維護、能源耗費、溫度監測體系搭建等相關支出，這部分成本隨溫度控制精度及運輸管理品質的提升而增加，要維持全程 $\pm 0.5^{\circ}\text{C}$ 的溫度穩定，需採用高性能製冷裝備與多點即時監測系統，所需投入遠高於 $\pm 2^{\circ}\text{C}$ 精度的溫度管控方案。

其二為損耗成本，即溫度變動引發的果實品質下滑、可售數量縮減及銷售價格走低造成的經濟損耗，損耗成本隨溫度控制精度的提升而下降，但下降態勢逐漸放緩，溫度控制精度達到特定水準後，繼續提升精度實現的損耗降低效果相對有限。

海南熱帶水果呼吸代謝旺盛，對溫度變動敏感度高，損耗成本曲線在溫度控制精度較低的區間呈現急劇下降特徵，溫度控制精度從 $\pm 5^{\circ}\text{C}$ 提升至 $\pm 2^{\circ}\text{C}$ 時，能明顯降低損耗比例，進一步將精度提升至 $\pm 1^{\circ}\text{C}$ 乃至 $\pm 0.5^{\circ}\text{C}$ ，雖可減少部分品質損失，額外增加的能源及裝備成本卻可能超出損耗減少帶來的經濟收益。

實際運輸決策階段，需結合水果品種特性、運輸里程、季節氣候及市場價格波動設定溫度控制精度，運輸高價值且耐儲能力弱的水果（如榴槤、山竹）時，可採用更高精度的溫度管控方案，降低高額的單果經濟損耗，對於耐儲性較強的品類（如椰子、香蕉），可在保障基礎品質的前提下選取成本較低的溫度管控標準，達成整體經濟收益最大化。

該理論在冷鏈管理中的有效應用，依賴於資料化監測及評估體系的支撐，通過記錄過往

運輸批次的溫度管控投入、溫度變動資料及實際損耗比例，可構建適配不同水果品類的成本—損耗關聯模型，運輸企業能夠基於模型預判不同溫度管控方案對應的經濟成效，選定最優管理策略，海南熱帶水果的出口與內銷業務中，這類平衡決策有助於降低運輸環節的綜合成本，穩固市場競爭力的同時提高產業整體盈利水準。

3.4 典型案例分析

基於海南熱帶水果冷鏈運輸中溫度變動與貨損比例關係的理論分析，引入具體實例可更直觀展現溫度管控在不同運輸場景下對果實品質與損耗的實際作用，本節選取三類具有代表性的運輸模式，包括出口物流、內銷長途運輸及短途區域配送，通過案例探析溫度變動在不同條件下的表現特徵及管理啟示。

案例一：海南金煌芒出口新加坡的海運物流

某出口企業 2022 年夏季向新加坡批量出口金煌芒，運輸里程約 1900 公里，全程採用海運冷藏集裝箱，依據出口商提供的溫度記錄設備資料，前 3 天運輸溫度穩定維持在 13°C ，裝卸港口等候期間，製冷裝置短暫故障，集裝箱內溫度升至 18°C 並持續約 7 小時，抵達新加坡後，檢驗發現約 12% 的芒果出現果皮發黑、果肉過早軟化現象，市場銷售價格較預期下降 18%，經後續核算，溫度變動造成的經濟損失高達 4.6 萬元人民幣，該案例顯示，短時溫度異常會對熱帶水果出口品質產生顯著影響，果實臨近呼吸躍變高峰期時，溫度變動會加速成熟與腐敗進程。

案例二：海南鳳梨北運北京的公路物流

2023 年春季，一批海南鳳梨通過公路運輸發往北京，全程約 2800 公里，運輸時長約 3.5 天，承運商為壓縮運輸成本，將溫度控制精度設定於 $\pm 3^{\circ}\text{C}$ 範圍，運輸途中，受北方晝夜溫差大影響，車廂內溫度夜間最低降至 5°C ，白天最高升至 12°C ，到達目的地後，批發市場檢測發現部分鳳梨出現低溫傷害（表皮出現褐色斑紋）且糖度降低，整體損耗比例為 8.5%，與以往採用 $\pm 1^{\circ}\text{C}$ 精度運輸的批次相比，雖節省約 15% 的能源費用，損耗增加帶來的直接經濟損耗超過節省成本的兩倍，此案例凸顯長途運輸中忽

視溫度管控穩定性對果實品質的負面影響，為成本 — 損耗平衡提供實際參考依據。

案例三：海南火龍果本島內短途配送的溫度管控疏漏

海南本島內，某水果批發商從種植基地向海口市內商超配送火龍果，全程不足 100 公里，運輸時長約 2 小時，因運輸路程較短，承運方未啟動冷鏈裝置，借助凌晨低溫環境實施自然運輸，當日氣溫驟升至 32℃，車廂內溫度在運輸後半程快速接近外界溫度，導致果實到貨後 24 小時內出現失水比例上升、果皮色澤暗沉等品質劣化問題，造成約 5% 的貨損比例，損失比例不大，但火龍果單價較高，直接影響商超的銷售利潤，該案例表明，短途運輸中溫度變動仍可能引發明顯品質變化，高溫季節，缺乏基礎溫度管控措施會放大風險隱患。

這三個實例共同表明，溫度變動的影響與運輸距離、運輸時長相關，還與運輸環節的溫度管控策略、果實成熟程度、季節環境條件密切相關，對於海南熱帶水果，構建全程溫度管控監測體系、制定品種差異化的溫度控制規範，在關鍵環節（裝卸作業、等候期間、中轉運輸）實施風險防控，是降低貨損比例、穩定果實品質的核心舉措。

4 國內外冷鏈運輸溫控管理經驗借鑒

4.1 發達國家的溫控經驗

全球冷鏈運輸領域，發達國家的溫度管控模式始終保持領先水準，其制度設計、技術應用與管理方法在控制溫度變動、減少貨損比例方面成效突出，美國、日本及歐盟各成員國等，在熱帶水果及其他易腐農產品的運輸環節，廣泛構建“全程可控、全程可溯”的溫度管控體系，這些實踐經驗體現在硬體設施的先進配置，也反映在制度執行力度、行業標準化程度及資訊化技術的深度融合。

美國冷鏈運輸環節由多部門協同監管，運輸過程中配置精度可達 $\pm 0.5^{\circ}\text{C}$ 的電子測溫記錄設備，借助物聯網系統完成資料即時上傳與動態

監控，溫度超出限定範圍，系統會自動向承運企業與貨主發送預警資訊，還可在車輛行駛過程中通過遠端操作調節製冷設備功率，行業內廣泛推行“產地預冷 — 恆溫運輸 — 快速配送”的三段式溫度管控方案，產地預冷階段通過強制空氣冷卻快速將果溫降至適宜區間，恆溫運輸階段保持溫度變動不超過 $\pm 0.5^{\circ}\text{C}$ ，大幅延長熱帶水果保鮮週期，減少運輸途中的生理病害。

日本冷鏈體系以精細管控與運行穩定為突出特點，在熱帶水果運輸中嚴格把控溫度，同步調控相對濕度與氣體構成，針對芒果、鳳梨等水果採用溫濕雙控模式，結合氣調技術通過調低氧氣占比、提高二氧化碳占比抑制果實呼吸作用，延緩成熟與衰老進程，日本港口冷藏裝備與運輸車輛達成無縫對接，裝卸操作中果實基本不暴露在常溫環境，顯著降低裝卸環節的溫度變動，行業協會定期對冷鏈企業開展分級評定，評定結果直接關係企業市場准入資格與合作資源，這種制度化的行業競爭環境有效提升整體服務品質。

歐盟國家在溫度管控中凸顯法規統一化與跨境協同優勢，明確各類水果的適宜運輸溫度範圍及允許變動幅度（通常為 $\pm 1^{\circ}\text{C}$ ），運輸過程中普遍採用多點測溫監測技術，在集裝箱或冷藏車內佈設多個測溫探頭，捕捉空間內可能存在的局部溫差，這些資料經由雲服務平臺向貨主、承運方及監管部門共用，實現全程透明化與可追溯性，近年歐盟探索將區塊鏈技術融入冷鏈管控，使溫度資料、防篡改記錄與責任追溯有機結合，大幅提升溫控資料的可信度與管理效率。

綜合來看，發達國家冷鏈運輸溫控管理的優勢集中體現在四個方面：一是法規標準的剛性約束作用，確保溫度管理有章可循、執行到位；二是高精度溫度管控技術的普及推廣，降低溫度變動風險；三是全程可溯的資訊化管理體系，提升透明度與管理回應效率；四是行業自律機制與市場競爭環境，推動冷鏈企業持續優化服務水準。

這些實踐經驗對海南熱帶水果冷鏈運輸有著重要參考價值，海南在硬體設施建設、法規政策

配套、資訊化管控及行業競爭機制建設方面協同發力，有望顯著降低運輸過程中的溫度變動幅度，減少果實品質劣變與貨損比例，為熱帶水果產業的品牌建設與國際化發展築牢基礎。

4.2 熱帶水果出口國的溫控模式

全球熱帶水果供應體系裡，泰國、菲律賓、越南等出口國依託豐富的熱帶水果品類儲備及長期國際貿易實踐，構建起成熟度較高且兼具地域特色的冷鏈溫度管控模式，這些國家的溫度管控在硬體水準與資訊化程度上不及部分發達國家，在應對高溫高濕環境、長距離跨境運輸及成本管控等方面積累了獨特實踐經驗。

泰國是全球重要的熱帶水果出口地，其芒果、榴槤、龍眼等產品出口規模位居世界前列，當地冷鏈架構注重“產地初加工與預冷處理”的關鍵作用，多採用水冷、真空預冷等技術在採收後兩小時內快速降低果實溫度，減少呼吸強度與水分蒸發，運輸過程中，泰國企業普遍使用隔熱效果優良的集裝箱，搭配冰塊、冰膠包等輔助控溫措施，應對長途運輸中可能出現的製冷設備波動或能源中斷問題，針對出口中國、韓國等市場的需求，泰國部分大型出口商在集裝箱內加裝多點溫濕度監控系統，確保運輸過程中的環境參數符合進口國檢驗檢疫要求。

菲律賓在香蕉、鳳梨等大宗熱帶水果出口中採用“單品專線”運作模式，即為單一品類水果制定專屬冷鏈運輸規範與操作流程，例如出口香蕉採收後會立即進入分級、清洗及浸泡抗菌液處理步驟，隨後轉入溫度約 13℃ 的預冷間進行梯度降溫處理，避免“冷害”現象發生，運輸階段維持恒溫狀態，變動範圍通常不超過 $\pm 1^{\circ}\text{C}$ ，貨櫃內設置空氣迴圈裝置，減少局部溫差引發的品質差異，菲律賓還廣泛推行“港口直裝”模式，縮短貨物在港口的堆存時長，降低裝卸過程中的溫度變動風險。

越南近年在火龍果、芒果、荔枝等出口業務中逐步搭建區域性冷鏈集散樞紐，吸納部分發達國家的溫度管控技術，與泰國、菲律賓不同，越南更注重與進口國標準的銜接，部分出口企業在集裝箱冷藏設備中配備遠端資料傳送功能，使買方可即時監控溫度曲線，這種透明

化管控模式在提升客戶信任度、減少貿易糾紛方面發揮積極作用，受冷鏈基礎設施整體薄弱、運輸裝備更新進程遲緩影響，越南長距離運輸中仍存在溫度變動較大的問題，一定程度上影響果實抵達目的地時的外觀與口感。

熱帶水果出口國在溫度管控中形成的共性經驗包括：高度重視採後預冷工作，根據品類特性設定專屬運輸標準，結合經濟條件採用靈活溫控措施，在關鍵環節盡可能縮短常溫暴露時長，這些經驗雖在技術先進性上不及發達國家，但在應對高溫環境、成本約束及長途跨境運輸實際條件方面具有較強適應性，對於海南而言，這些國家的做法在採後處理、分品類溫度管控及經濟性方案設計等方面具備一定參考價值，尤其適用於當前海南熱帶水果出口規模持續擴大的背景下，探尋有限投資條件下提升溫度管控水準的路徑。

4.3 對海南的可借鑒要點與局限性

對照泰國、菲律賓、越南等熱帶水果出口國的溫度管控實踐，這些經驗在海南熱帶水果冷鏈運輸體系建設與優化中具備一定參考價值，但引入應用過程中存在適用局限，需要結合本地實際進行篩選與調整。

可借鑒要點方面，採後快速預冷是降低溫度變動、延緩果實衰老的核心步驟，泰國採後兩小時內完成預冷處理的做法值得海南推廣，尤其是針對芒果、火龍果、鳳梨等溫度敏感度較高的品類，通過搭建產地型預冷樞紐、優化冷鏈節點排布，可顯著降低運輸過程中溫度變動的初始風險，分品類定制溫度管控模式有助於減少品質損失，例如菲律賓為香蕉、鳳梨設定精確穩定的運輸溫濕度標準，海南可在龍眼、荔枝、芒果等品類上分別建立相應標準化運輸參數，並納入行業指導性檔，縮短常溫暴露時長是降低貨損比例的關鍵舉措，菲律賓的“港口直裝”與越南的“冷鏈集散樞紐”模式均能減少裝卸環節的溫度變動，這對海南尤為重要，海南水果外運需經港口或冷鏈中轉站，優化裝卸流程可有效減少溫度管控中斷的機會。

局限性方面，海南在冷鏈基礎設施佈局與裝備更新方面仍有短板，尤其是部分農村產區

的預冷及儲藏條件不足，直接複製泰國或菲律賓的高頻預冷模式可能面臨成本與設備保障欠缺的困難，海南熱帶水果的銷售管道與運輸里程呈現多元化特點，部分為短途內銷，部分為長途北運或出口，這種多元化運輸需求意味著單一溫度管控模式難以全面適配所有場景，需結合運輸半徑、品類特性與市場要求制定靈活溫度管控策略，海南冷鏈資訊化管控的整體水準尚有提升空間，越南出口企業常用的遠端溫度監控系統雖能提升透明度，但其在海南的全面推廣可能受到運營成本、技術培訓及終端接受度的限制。

海南借鑒國外經驗時，應在技術引進、模式應用與制度保障三個層面進行本土化適配，技術上可分階段導入遠端監控與多點溫濕度檢測裝備，模式上優先推廣“產地預冷 + 直達運輸”的組合架構以降低中間環節風險，制度上可通過地方行業協會擬定品類溫度管控指南，推動港口與物流企業協同優化裝卸環節，這種有選擇、有針對性的借鑒下，國外經驗才可能在海南熱帶水果冷鏈運輸體系中落地並發揮實效。

5 海南熱帶水果冷鏈運輸溫控優化建議

基於前文系統分析，海南熱帶水果冷鏈運輸的溫控優化需以減小溫度變動幅度、延緩品質劣變進程、降低貨損比例為核心導向，結合本地氣候特徵、產業佈局及市場訴求，搭建從產地源頭到終端市場的全鏈條管控方案。

其一，優化產地預冷體系與冷鏈基礎設施配置，海南熱帶氣候使水果採收後易出現呼吸作用加劇、水分快速流失問題，高效預冷成為抑制溫度變動的首要屏障，建議在主要水果產區搭建集中式預冷樞紐，配備真空預冷、強制風冷等多元預冷技術，適配不同品類的特性需求，通過財政補貼與產業扶持政策，引導合作社與龍頭企業協作搭建中小型預冷網站，覆蓋偏遠產區，減少果實採收後長時間暴露在常溫環境中的風險。

其二，構建分品類溫控規範與動態適配機

制，海南水果品類豐富，不同果品對溫度、濕度的適配範圍存在明顯差異，應依託科研機構與行業協會，制定涵蓋芒果、荔枝、火龍果、鳳梨、龍眼等主力品類的運輸溫控參數，將其納入行業推薦規範或地方技術標準，實際運輸過程中，需配備多點溫濕度記錄設備，結合運輸里程、季節氣溫波動及貨物堆碼密度，動態調整製冷強度與通風頻次，避免單一溫控設定引發的能耗浪費或品質損傷。

其三，提升運輸與裝卸環節的協同效能，海南水果外運常涉及港口或冷鏈中轉網站，裝卸過程中的溫控中斷是誘發溫度變動的關鍵因素，建議推動港口、物流園區與果品企業建立裝卸預約機制，縮短貨物露天等候時間，優先配置封閉式裝卸通道或移動冷庫，推廣“產地直達 — 終端市場”的一體化冷鏈運輸模式，減少中途多次裝卸流程，降低溫度失控的潛在風險。

其四，強化冷鏈運輸的資訊化與智慧化管控能力，通過引入遠端溫控監測系統與物聯網溫濕度傳感設備，實現運輸全程的即時監控與異常預警，貨物抵達後生成完整溫控報告，作為品質溯源與責任界定的依據，這一舉措能提升物流環節的透明度，在貨損發生時為保險理賠提供資料支撐，針對中小運輸企業，可由行業協會或政府部門牽頭搭建共用型溫控監測平臺，降低技術應用的門檻與運營成本。

其五，建立多方協同的溫控管理體系，冷鏈運輸涵蓋種植、加工、物流、銷售等多個環節，單一主體難以實現全程溫控的穩定性，建議在海南省內探索“政府引導 + 行業規範 + 企業自律”的協同模式，由政府出臺冷鏈溫控管理指導檔，行業協會負責標準推廣與專業培訓，企業在執行過程中結合實際開展技術創新與成本優化，設立冷鏈運輸品質評價與激勵機制，對溫控表現優異的企業給予政策扶持與市場推廣資源，形成良性競爭與持續改進的行業氛圍。

通過上述優化策略的統籌實施，海南熱帶水果冷鏈運輸中的溫度變動有望得到有效管控，果實品質與貨損比例將得到顯著改善，能夠提升本地水果在國內外市場的競爭實力，為

全省熱帶農業的高品質發展築牢基礎。

6 對海南水果供應鏈的綜合影響

冷鏈運輸溫度管控優化直接作用於水果運輸過程中的品質與損耗程度，對海南水果供應鏈的結構、運作效能和市場表現產生系統性作用^[7]，溫度波動控制能力的提升，意味著從產地到消費者的各個環節都將獲得更高的穩定性與可預期性，這種變化將通過以下方面體現。

其一，對上游種植端而言，穩定的溫度管控能夠明顯減少採後損耗隱患，使種植戶在銷售與議價過程中擁有更足的主導權，以往冷鏈穩定性不足，部分農戶在銷售高峰期不得不以偏低價格迅速銷售，避免品質下滑造成的損失，溫度管控改善後，農戶可延長銷售週期，選擇更有利的市場時機，進而提升收入水準，品質穩定會推動種植者更注重優質化種植方式，形成從種植到運輸的正向激勵效應。

其二，對中游物流與批發環節而言，溫度管控優化提高了運輸的可預期性與合格比例，降低了因品質爭議產生的退貨及理賠矛盾，對批發商來說，貨物到港後的品相一致性增強，有助於快速完成分銷，減少滯留與二次損耗，在港口、冷鏈中轉站等關鍵節點，溫度管控穩定性還能減少中途篩貨、挑揀的時間成本，提升作業效能，增強與大型零售商及出口客戶的合作信任度。

其三，對下游零售端與消費者而言，溫度管控改善將直接改善終端水果的口感、品相與保鮮週期，進而提高消費者對海南水果品牌的信賴度與再次購買率，尤其在高端市場與出口市場，品質穩定性是維持價格優勢與品牌溢價的核心因素，溫度管控優化有助於塑造海南熱帶水果在國內外市場的高端定位，推動從“原產地標識”向“品質保障標識”的轉變。

此外，溫度管控改善還將促進海南水果供應鏈的資訊化建設與資料化管理，溫度管控資料的即時監測與留存，有助於追溯貨損成因，為整個供應鏈提供決策依據，例如分析不同運輸路線與季節的溫度管控資料，可優化運輸路線與調度計畫，降低整體物流成本，長期積累的資料還可成為政府與行業制定冷鏈標準及政

策的基礎，提高供應鏈治理的科學水準。

更為重要的是，溫度管控優化將在一定程度上推動海南水果產業鏈的縱向整合與橫向聯動，龍頭企業在掌握穩定溫度管控能力後，更有動力整合上游種植資源與下游銷售管道，形成“產運銷融合化”模式，進而提升整個供應鏈的議價能力與抗風險能力，溫度管控優化還會促使更多中小企業通過合作社、產業聯盟等方式實現資源共用，降低設備投資與技術應用的門檻，推動行業整體水準的提升。

總體來看，冷鏈運輸溫度管控的優化能降低貨損比例與運輸損耗，更將通過改善品質穩定性、提升市場競爭力與優化供應鏈協作模式，為海南熱帶水果產業的可持續發展提供堅實支撐，這種影響是結構性、長期性的，既體現在經濟效益的提升，也反映在品牌建設與產業升級的進程之中。

7 結論

海南熱帶水果在國內外市場的競爭力在很大程度上依賴於冷鏈運輸過程中品質的維持與損耗的管控，溫度波動作為影響果實品質與貨損比例的核心因素，其管理水準直接決定了產品從產地到終端的價值實現程度，通過對海南冷鏈運輸現狀、溫度波動影響機理、理論分析框架以及國內外溫度管控經驗的系統探討，可得出以下結論。

第一，海南熱帶水果冷鏈運輸在基礎設施、溫度管控技術與管理模式方面已獲得一定發展，但溫度波動仍然普遍存在，尤其在長途運輸、節點轉運及末端配送環節，波動幅度較大，對果實品質造成不可逆轉的損傷，提高了貨損比例。

第二，溫度波動對果實品質的影響具有生物學與物理學雙重作用機制，涉及呼吸作用增強、乙烯釋放量上升、組織失水、病害發生等多個過程，這些變化在不同水果品種、成熟度與運輸環境下表現差異顯著，體現出溫度管控必須具備針對性與精細性的特點。

第三，貨損比例的變化不僅取決於溫度波動幅度本身，還受運輸時長、包裝方式、品種耐

貯性以及管理回應速度等因素的協同影響，成本 — 損耗平衡理論表明，在溫度管控投入與損耗減少之間存在最佳均衡狀態，合理配置資源可明顯提高經濟效益與品質穩定性。

第四，發達國家與熱帶水果出口國在溫度管控方面的經驗，為海南提供了可參考的技術與管理思路，如全過程溫度管控監測、動態調溫、分區控溫、即時資料共用等，海南在氣候條件、產業結構與市場需求方面具有獨特性，借鑒模式必須經過本土化調整。

第五，溫度管控優化不僅是降低貨損的技術手段，更是推動海南熱帶水果供應鏈升級的核心環節，其改善效應將通過品質提升、市場

拓展、品牌塑造與產業協同等途徑，帶來結構性、長期性的競爭優勢。

綜上，海南熱帶水果冷鏈運輸的溫度波動問題亟需在技術、管理與制度層面協同應對，未來應通過加強基礎設施建設、引入智慧化溫度管控設備、建立全過程溫度追溯體系、推動供應鏈各主體的協同合作，構建高效穩定且可持續的冷鏈運輸體系，唯有如此，才能在激烈的市場競爭中穩固海南熱帶水果的品牌地位，實現產業的高品質發展。

參考文獻：

- [1] iResearch 諮詢院. 中國生鮮農產品供應鏈研究報告-2020年[EB/OL]. 2020-03-31.
- [2] 鄭小瓊. 海南省熱帶水果冷鏈物流發展策略探析[J]. 現代商業, 2016, (16): 21-22. DOI:10.3969/j.issn.1673-5889.2016.16.007.
- [3] Nwokocha B C, Chatzifragkou A, Fagan C C. Impact of ultrasonication on African oil bean (*Pentaclethra macrophylla* Benth) protein extraction and properties[J]. Foods, 2023, 12(8): 1627. DOI:10.3390/foods12081627.
- [4] Wu J, Tang R, Fan K. Recent advances in postharvest technologies for reducing chilling injury symptoms of fruits and vegetables: A review[J]. Food Chemistry: X, 2023, 21: 101080. DOI:10.1016/j.fochx.2023.101080.
- [5] Feng Q, Wang Z, Xiong W, et al. The effect of postharvest storage temperatures on fruit flavor constituents in ‘Wushancuili’ plum[J]. Horticulturae, 2024, 10(4): 414. DOI:10.3390/horticulturae10040414.
- [6] Zehra K, Nawab A, Alam F, et al. Development of novel biodegradable water chestnut starch/PVA composite film: Evaluation of plasticizer effect over physical, barrier, and mechanical properties[J]. Journal of Food Processing and Preservation, 2022.
- [7] Moffat A M, Beckstein C, Churkina G, et al. Characterization of ecosystem responses to climatic controls using artificial neural network[J]. Global Change Biology, 2010, 16(10): 2737-2749.

版權聲明

© 2025 作者版權所有。本文依據“知識共用署名 4.0 國際授權合約”（CC BY 4.0）以開放獲取方式發佈。該許可允許使用者在任何媒介中自由使用、複製、傳播與改編文章（含商業用途），惟須明確署名原作者及出處，並注明所作修改（如有）。完整協議詳見：<https://creativecommons.org/licenses/by/4.0/deed.zh-hans>

出版聲明

所有出版物中的陳述、觀點及資料僅代表作者及供稿者個人立場，與 Brilliance Publishing Limited 及/或編

輯人員無關。Brilliance Publishing Limited 及/或編輯人員對因內容所提及的任何理念、方法、說明或產品所導致的人身或財產損害概不負責。

DOI: 10.53104/xdkxsts.2025.01.02.002

AIGC 生成內容的權威性與品質控制科學知識生成中的驗證與評價 機制探討

陳明¹

1. 南京師範大學，江蘇 南京，210023

摘要：人工智慧生成內容（AIGC）技術持續高速演進推動知識生產與傳播模式發生深度轉型，AIGC 在文本創制、圖像生成、資料解析與科研協助等多領域落地大幅提升資訊處理效能與知識創造速率，AIGC 在內容品質與權威特質方面的潛在隱患逐步顯現，科學知識生產與公共傳播領域中其內容真實屬性、可靠特質與可驗證屬性等問題亟需回應。本文系統梳理 AIGC 的技術發展背景與應用現狀從內容品質管控、權威特質驗證與評價機制構建三個維度解析其核心技術邏輯與制度約束條件，研究表明當前 AIGC 內容在真實屬性評估、倫理責任界定與標準化治理方面仍存在明顯缺口集中表現為事實核驗機制不足、演算法偏見積澱與統一品質評估架構缺失。文章進一步探析 AIGC 內容權威性的社會認知基礎與科學驗證機制的結構性難題提出通過搭建可解釋性演算法體系、健全多層次驗證機制、建立跨行業品質基準與倫理治理架構優化 AIGC 知識生成的可信程度，本文認為 AIGC 的未來發展取向應聚焦“科學驗證 — 品質評估 — 社會信任”三維聯動推動人工智慧從“資訊生成”邁向“知識生產”實現技術創新與認知可靠屬性的動態平衡。

關鍵字：人工智慧生成內容；權威性；品質控制；驗證機制；知識生成；可信 AI

Exploring the Authority and Quality Control of AIGC Generated Content in the Verification and Evaluation Mechanisms of Scientific Knowledge Production

CHEN Ming¹

1. Nanjing Normal University, Nanjing 210023, P.R.China

Correspondence to: CHEN Ming; Email: chen_ming74@hotmail.com

Abstract: The rapid evolution of Artificial Intelligence Generated Content (AIGC) technology has profoundly transformed the modes of knowledge production and dissemination. The application of AIGC across multiple domains—including text creation, image generation, data analysis, and scientific research assistance—has greatly enhanced the efficiency of information processing and the speed of knowledge generation. However, potential risks concerning content quality and authority have gradually emerged, particularly in the fields of scientific knowledge production and public communication, where issues such as factual authenticity, reliability,

收稿日期：2025-11-19 返修日期：2025-12-16 錄用日期：2025-12-22 出版日期：2025-12-31

通信作者：chen_ming74@hotmail.com

引用格式：陳明. AIGC 生成內容的權威性與品質控制科學知識生成中的驗證與評價機制探討[J]. 現代科學探索, 2025, 1(2): 12-29.

and verifiability of AIGC-generated content demand urgent attention. This paper systematically reviews the technological background and application status of AIGC, analyzing its core technical logic and institutional constraints from three dimensions: content quality control, authority verification, and evaluation mechanism construction. The study finds that current AIGC content still faces significant deficiencies in factual verification, ethical accountability, and standardized governance—mainly reflected in insufficient fact-checking systems, accumulated algorithmic bias, and the lack of unified quality assessment frameworks. Furthermore, the paper explores the social cognitive foundations of AIGC authority and the structural challenges of scientific validation mechanisms. It proposes the construction of explainable algorithmic systems, multi-layered verification frameworks, and cross-industry ethical governance standards to enhance the credibility of AI-generated knowledge. The study concludes that future development of AIGC should focus on a triadic linkage among “scientific validation – quality assessment – social trust,” promoting a transition of artificial intelligence from “information generation” to “knowledge production,” thereby achieving a dynamic balance between technological innovation and epistemic reliability.

Key words: Artificial Intelligence Generated Content (AIGC); authority; quality control; validation mechanism; knowledge generation; trustworthy AI

引言

人工智慧技術高速發展讓生成式人工智慧內容 (AIGC, AI-Generated Content) 成為當下社會及各行業的重要組成部分，這項融合深度學習、自然語言處理、電腦視覺等技術的內容生成方式能自主產出文章、新聞稿件、廣告構想、學術論著、圖像、音訊及視頻等多種形式成果，為各行各業提供極大便利的同時提升生產效能、降低人力成本，在個性化推薦、內容創作等領域已成為核心驅動力量。

AIGC 技術應用不斷深入使其生成內容的品質與權威性受到廣泛關注，依賴海量歷史資料訓練的 AIGC 模型因資料品質與來源難以完全可靠，導致生成內容在精準度、權威性、客觀性等方面存在潛在風險，且這類內容會受演算法模型自身局限性、輸入資料品質、模型偏差及外部環境變動等多重因素影響，保障 AIGC 內容生成效率的同時確保其品質與可信度成為技術界、學術界及行業發展的重大難題。

AIGC 生成內容的權威性問題尤為突出，資訊傳播節奏加快的當下網路資訊的真實性與精準度面臨前所未有的檢驗，缺乏有效品質管控與驗證機制的 AIGC 內容可能成為誤導公眾、傳播不實資訊的工具並影響社會決策、輿

論走向及公眾信任，而 AIGC 生成內容的評價標準與方法仍處於探索階段，現有品質評判體系與驗證機制未能完全適配這一快速發展的技術領域，建立健全 AIGC 內容品質管控與驗證體系成為保障其各領域健康發展、避免技術濫用引發社會問題的迫切需求。

本文基於 AIGC 生成內容的技術背景與應用現狀，探討其品質管控與驗證機制的現有狀況及面臨的挑戰，具體將圍繞以下方面展開：回顧 AIGC 技術的起源與發展歷程，分析其核心技術架構與工作機理，探究當前 AIGC 技術在各行各業的應用趨勢及主要挑戰；剖析 AIGC 生成內容的品質管控機制，探討現有品質評判標準與管控手段，指出當前品質管控體系的薄弱環節並提出未來優化方向；重點解析 AIGC 生成內容的權威性與可信度，明確二者定義，探究影響 AIGC 內容權威性的各類因素並提出提升內容可信度的策略與實踐路徑；詳細探討 AIGC 內容的驗證與評價機制，分析現有技術與方法的挑戰及創新方向，探究現有驗證體系的優化路徑；總結當前研究成果並提出相關政策建議，對 AIGC 技術未來發展方向進行展望。

AIGC 的快速發展帶來極大便利的同時伴隨不可忽視的風險，全球資訊化進程持續提速背景下有效保障 AIGC 生成內容的品質、權威

性及社會價值成為技術發展無法回避的重要課題，本文旨在為這一課題的解決提供理論支撐與實踐框架，推動 AIGC 相關領域的深入研究与技術創新，為政策制定者提供參考以助力 AIGC 技術健康發展。

本研究期望為完善 AIGC 生成內容的品質管控與驗證機制提供有力支撐，增進社會對 AIGC 技術的正確認知與合理運用，確保其在更健康、更可控的環境中發揮更大社會效益。

1 AIGC 生成內容的技術背景與現狀

1.1 AIGC 技術的起源與發展

AIGC（人工智慧生成內容）作為新興技術領域依託人工智慧技術演進形成，深度學習、自然語言處理與電腦視覺的快速發展為其提供核心支撐，通過演算法模型自主產出與人類創作成果相近的文本、圖像、視頻等內容，脫離傳統人工創作流程的約束，背後基礎理論與實踐應用已通過長期反覆運算實現逐步完善^[1]。

AIGC 技術的雛形源自人工智慧早期探索階段，20 世紀 50 至 80 年代 AI 領域研究集中於邏輯推演與符號運算，這一時期的電腦程式僅用於執行特定規則與任務不具備創意發揮或內容生成能力，90 年代機器學習技術興起讓 AI 逐步獲得學習與資料處理能力進入“弱人工智慧”階段，得以在特定領域內完成指定工作。

21 世紀人工神經網路與深度學習的崛起為 AIGC 技術發展奠定堅實基礎，卷積神經網路（CNN）、遞迴神經網路（RNN）及生成對抗網路（GAN）等深度學習模型實現 AIGC 技術的革命性突破，電腦借助深度神經網路從海量資料中捕捉複雜模式並開展自我訓練，最終具備自主生成內容的能力。

2014 年 Ian Goodfellow 團隊提出的生成對抗網路（GAN）成為 AIGC 技術發展的關鍵節點，該模型通過對抗性訓練模式讓生成模型更精準類比真實資料的分佈特徵，使其在圖像、視頻及文本生成等領域展現出前所未有的能力，既能產出高度逼真的圖片也能依據少量輸

入創作富有創意的藝術作品、廣告文案等內容。

GAN 的成功推動圖像生成領域進步的同時激發文本生成領域的研究熱潮，基於 Transformer 架構的生成模型（如生成預訓練 Transformer GPT）實現自然語言處理技術的重大突破，GPT 系列模型通過預訓練與微調的結合生成語法規範、語義豐富的自然語言文本，確立在文本生成領域的標杆地位^[2]。

OpenAI 推出的 GPT-1、GPT-2 及 GPT-3 系列生成預訓練語言模型在自然語言生成任務中表現卓越，其中 GPT-3 以 1750 億參數規模引發學術界與工業界廣泛關注，可完成文本創作、代碼編寫、對話生成、詩歌創作等多項任務，既提升 AIGC 在文本內容生成領域的品質水準也拓展其應用邊界。

GPT-3 的普及應用讓 AIGC 技術影響力快速滲透至教育、醫療、金融、媒體、娛樂等多個行業，在自動寫作、智慧客服、情感分析、文本摘要等場景得到實際應用，圖像生成領域通過 DALL·E 等模型實現文本到圖像的順暢轉換，有效推動 AIGC 技術在實際場景中的落地實施。

深度學習與多工學習的發展讓 AIGC 應用突破單一資料類型限制，多模態生成技術成為 AIGC 領域的研究前沿，這類技術整合圖像、視頻、文本等多種資料形態實現跨模態資訊的生成與轉化，OpenAI 的 DALL·E 模型可依據文本描述創作高度擬真且富有創意的視覺作品，CLIP 模型實現視覺資訊與語言資訊的融合推動圖像理解與生成的跨領域協同。

多模態生成技術拓展 AIGC 在藝術創作、品牌行銷、娛樂內容製作等領域的應用範圍，技術升級帶來深度學習模型泛化能力的持續提升，AIGC 已能依據用戶輸入的少量提示或指令生成具備高度定制化與創新性的內容，使用者通過與 AI 的交互可快速獲取符合自身需求的創意成果顯著提升創作效率。

未來 AIGC 技術將朝著生成品質更優、應用場景更廣、個性化程度更高的方向持續演進，計算能力的提升與演算法的優化將推動 AIGC 在生成速度、內容多樣性與創意性方面實現更大突破，自然語言處理、電腦視覺與語音生成

等領域的融合應用將使其在更多新興行業中發揮重要作用。

AIGC 技術的普及應用讓保障生成內容的品質、準確性與可信度成為日益關鍵的議題，技術的快速發展既需要技術層面的持續革新也需要完善相應的品質控制、驗證機制及倫理法規，技術、標準與監管的協同發力才能推動 AIGC 走向成熟，為各行業創新發展提供堅實支撐。

AIGC 技術的起源與發展歷程映射人工智慧領域的迅猛進步，從早期簡單自動化任務到如今高複雜度內容生成，AIGC 已成為推動社會變革與行業發展的重要力量，技術的持續突破與應用場景的不斷拓展將讓其在未來內容生產與創作領域發揮愈發重要的作用。

1.2 AIGC 的核心技術架構與原理

AIGC（人工智慧生成內容）的核心技術架構與原理以多種前沿人工智慧方法及深度學習模型為支撐，生成對抗網路（GAN）、自回歸模型（如 GPT）、變分自編碼器（VAE）及多模態學習構成核心支撐體系，這些技術共同實現對人類創作類內容的理解與學習並生成形態相似的文本、圖像、音訊、視頻等，以下詳細拆解其核心技術架構與運作機理。

生成對抗網路（GAN）

生成對抗網路（GAN）是 AIGC 領域的核心技術支柱，2014 年由 Ian Goodfellow 提出，其結構包含生成器（Generator）與判別器（Discriminator）兩個核心神經網路，二者通過對抗性訓練形成博弈關係，生成器以產出與真實資料高度相似的內容為核心使命，判別器專注辨別生成內容與真實資料的差異。生成器通過持續學習資料分佈特徵從隨機雜訊中衍生圖像、文本等新的資料形態，判別器以區分資料真偽為目標判斷輸入樣本屬性，兩者在競爭中不斷優化生成器持續調整輸出形態直至判別器無法分辨其與真實內容的區別，這種機制讓 GAN 能夠生成高度逼真的各類內容並廣泛應用於圖像生成、文本創作及多模態內容生成等場景。

自回歸模型（如 GPT）

自回歸模型是 AIGC 技術體系中的重要架構，在自然語言處理（NLP）領域應用尤為廣泛，OpenAI 推出的 GPT 系列（Generative Pretrained Transformer）是該類模型中最具代表性的範例，其基於 Transformer 架構採用自回歸生成模式通過逐一生成詞彙或符號並借助已生成內容片段推測後續詞彙。GPT 模型先在海量的無標籤資料中完成預訓練掌握語言的統計特徵與上下文關聯，在無監督環境中積累廣泛語言知識後通過微調（Fine-tuning）適配文本生成、對話交互、情感研判等特定任務需求，依託對海量資料的學習積累 GPT 能夠生成連貫規範、語義豐富的自然語言文本並勝任文章創作、代碼編寫、語言翻譯等複雜語言類任務。

變分自編碼器（VAE）

變分自編碼器（VAE）是生成模型領域的常用技術，植根於概率圖模型的生成類框架，核心目標是通過學習資料潛在分佈特徵生成全新資料，其核心邏輯是編碼器將輸入資料轉換至潛在空間（Latent Space）維度解碼器再從這一空間還原原始資料特徵，訓練過程中 VAE 持續優化潛在空間分佈讓生成資料不斷逼近真實資料形態。與傳統自編碼器相比 VAE 引入概率模型設計通過變分推斷方式逼近真實資料分佈，進而產出更具真實性的資料樣本，在 AIGC 應用場景中 VAE 常被用於圖像生成、語音合成等領域且尤其適配需要多樣化內容輸出的使用場景。

Transformer 架構與多頭注意力機制

Transformer 架構在 AIGC 技術中具有不可替代的關鍵作用，處理文本、語音、時間序列等序列資料時優勢尤為突出，不同於傳統遞迴神經網路（RNN）與長短時記憶網路（LSTM），Transformer 借助“自注意力機制”（Self-Attention）並行處理序列中的全部元素顯著提升計算效率與模型表達能力。自注意力機制通過計算序列中每個元素與其他所有元素的關聯程度捕捉長距離依賴關係，彌補傳統 RNN 與 LSTM 處理長序列時的資訊損耗缺陷，Transformer 架構中的多頭注意力機制允許模型在同一層級並行關注輸入序列的多個子空間進一步強化模型表現力。在 AIGC 領域 Transformer 廣泛應用於文本生成、機器翻譯、

語音合成等任務，成為 GPT、BERT、T5 等大規模預訓練模型的核心基礎，其強大的處理能力支撐 AIGC 生成高品質文本內容並在多模態生成任務中發揮關鍵賦能作用。

多模態學習與生成

AI 技術的持續演進讓 AIGC 應用場景突破單一資料模態限制，多模態學習 (Multimodal Learning) 已成為當前 AIGC 領域的重點研究方向，這類學習模式指 AI 系統能夠同步解讀處理文本、圖像、音訊、視頻等多種不同模態的資料資訊並實現跨模態的資訊轉換及內容生成。OpenAI 的 DALL·E 模型可依據文本描述生成對應圖像，CLIP 模型則融合圖像與文本資訊完成圖像理解與搜索任務，多模態生成技術充分利用不同模態間的共性特徵與互補優勢，讓 AIGC 在廣告創意自動生成、視頻內容製作、跨語言翻譯等更多場景中實現高效創作與生產。

AIGC 的生成原理與演算法

AIGC 的生成過程可簡要歸納為模型通過訓練從海量真實資料中掌握資料的分佈特徵與規律再依託演算法完成內容生成，文本、圖像、音訊等各類生成任務的核心原理均為通過訓練讓模型在輸入文本、關鍵字、圖像描述等給定條件下產出滿足需求的目標內容。生成內容的品質與創意表現通常與訓練資料的品質及多樣性密切相關，通過持續優化訓練流程 AIGC 能夠生成具備高度創新性與複雜結構的內容，部分場景中 AIGC 還可借助生成對抗機制、強化學習等方式進一步提升生成內容的品質與多樣性水準。

AIGC 技術架構的整合

當前 AIGC 技術架構愈發重視模組化設計與跨領域融合，多個生成模型、學習演算法及資料來源可協同運作提升生成內容的品質與適用範圍，文本與圖像生成任務中，文本生成模型先輸出基礎文本內容再與圖像生成模型配合，通過 DALL·E 等技術生成契合文本描述的圖像，多模態 AI 平臺將各類技術整合銜接形成更為完整強大的 AIGC 技術架構體系，適配更廣泛的應用需求^[1]。

1.3 當前 AIGC 應用的行業趨勢與挑戰

人工智慧技術持續演進推動 AIGC (人工智慧生成內容) 在媒體、廣告、教育、醫療等多個行業深度落地，各領域均在主動探尋 AIGC 催生的行業變革，這類技術落地既提升生產運轉效率也給內容創作賦予更多創新方向與實現可能，AIGC 在實際場景應用中仍面臨不少現實難題，內容品質管控、倫理規範約束、版權歸屬界定等層面問題尤為突出，下文從行業應用趨勢與核心挑戰兩大維度展開深入剖析。

1.3.1 行業趨勢

內容創作領域的創新引領

AIGC 在內容創作領域滲透最為深入，新聞傳播、文學創作、廣告策劃等行業均有廣泛應用，GPT 系列、BERT 等自回歸模型已能自主完成新聞稿件撰寫、小說詩歌創作甚至科學研究論文框架草擬。廣告行業中 AIGC 被用於快速產出廣告文案、產品推介內容及行銷推廣方案，既提升創意從業者工作效能也有效控制人力投入成本。依託 GAN 技術的圖像生成平臺 (如 OpenAI 的 DALL·E) 可依據使用者文字描述產出高品質圖像，Deepfake 技術在影視製作、娛樂產業及虛擬實境 (VR) 領域獲得較多運用，借助 AIGC 相關技術打造虛擬角色與場景環境，技術持續升級將進一步推動個性化內容創作發展，更好滿足使用者對定制化內容的實際需求。

教育領域的智慧化革新

AIGC 在教育行業發展空間十分可觀，能為學習者提供定制化學習素材並自主生成練習題目、解題思路及學習評價，通過對話交互系統協助學生解決學習過程中各類問題，語言習得與程式設計教學場景中 AIGC 相關技術已能為學習者打造多樣化練習資源，助力學生依照自身學習節奏與個性化需求開展自主學習。AIGC 還被用於教育資源智慧化產出，自動構建線上課程體系、編制教學講義及製作教學視頻等，這種智慧化內容生產模式既提升教育資源可獲取程度也有效降低教師教學壓力，推動教育事業朝著普及化與智慧化方向穩步發展。

金融領域的智慧分析與報告編制

金融行業中 AIGC 核心應用集中在資料解讀、報告編制與智慧客戶服務等場景，能自

動生成財務分析報表、市場動態研判報告及投資參考建議，幫助金融從業者高效解析海量資料資源並為客戶提供貼合個人需求的投資規劃方案。AIGC 可借助自然語言生成技術（NLG）搭建智慧問答平臺，優化客戶服務回應速度並減輕人工客服工作負荷。

醫療行業的輔助診斷與健康養護

醫療領域中 AIGC 應用重點集中在醫學影像解讀、病歷概要撰寫、健康管理及定制化醫療服務等方面，通過大規模醫療資料訓練學習 AIGC 相關技術能夠協助醫生生成病例分析報告，為快速確診病症提供輔助支援，醫學影像領域 AIGC 已廣泛用於病灶識別、圖像標記及報告自動生成等工作，大幅提升醫學影像分析效率與精準度。AIGC 在健康管理領域也發揮重要作用，通過生成定制化健康指導方案與飲食搭配建議為用戶提供專屬健康養護規劃，AI 技術不斷進步讓 AIGC 在醫療行業應用前景更為廣闊，人口老齡化社會背景下其技術優勢將得到更為充分的展現。

1.3.2 面臨的挑戰

內容品質管控與偏見規避難題

AIGC 在內容創作方面展現突出應用價值但生成內容的品質水準與精準程度仍存在較高不確定性，部分文本生成模型因訓練資料存在偏向性，產出內容可能包含錯誤資訊、歧視性表述、片面觀點或資訊失實等問題，部分 AI 寫作工具生成的文章可能出現誇大其詞或虛假表述的情況，涉及敏感議題時極易引發倫理層面爭議。解決這一問題需進一步優化 AIGC 模型訓練資料資源，保障資料多元性與公正性，建立健全品質管控機制與驗證流程，確保 AIGC 生成內容符合真實客觀基本準則。

倫理規範與法律界定問題

AIGC 技術落地應用伴隨諸多倫理與法律層面挑戰，生成內容的版權所屬問題已成為行業內熱議焦點，這類內容版權究竟歸屬於誰、內容創作者與 AI 技術開發者之間版權劃分如何界定，目前尚未形成統一界定標準。Deepfake 等技術的不當使用還引發涉及個人隱私、虛假資訊擴散及肖像權侵害等方面的法律與倫理爭議，解決這些問題需要制定針對性法律法規條

文，強化對 AIGC 生成內容的監管力度，將 AIGC 生成內容的來源追溯、責任認定及隱私保護等事項納入法律約束框架，推動行業朝著合規化方向發展^[4]。

技術模型的可靠性能與透明程度

當前 AIGC 技術雖實現長足發展但多數模型的透明程度與可解釋性均顯不足，用戶難以清晰知曉 AIGC 生成內容的決策邏輯與過程，進而影響 AIGC 應用的可信度與市場接受度，醫療、金融等高風險行業中 AIGC 的“黑箱”特性可能引發決策失誤並造成嚴重不良後果。提升 AIGC 技術可靠性能與透明程度需要未來更加注重模型可解釋性研發，打造能夠向使用者呈現內容生成過程與決策依據的相關工具，同時增強演算法穩定性能避免生成錯誤或有害的內容資訊。

計算資源佔用與能源消耗問題

AIGC 相關技術尤其是基於深度學習的生成模型（如 GPT-3、DALL·E）的訓練與推理過程需耗費海量計算資源，這類大規模模型訓練通常需要高端計算硬體支援且消耗大量能源，帶來沉重的生態負擔與經濟開銷，保障技術性能前提下減少計算資源佔用成為 AIGC 技術發展過程中面臨的重要難題。技術創新可改良模型結構設計提升計算處理效能，推進綠色算力技術發展降低 AIGC 應用對環境造成的影響，這也是行業當前亟待解決的問題。

2 AIGC 生成內容的品質控制機制

本章聚焦 AIGC 生成內容在技術與流程維度的品質管控議題，重點探討優質內容的界定標準及現有系統在實際運作中通過哪些技術與管理路徑維繫內容的穩定性與可信度。本章關注的核心是模型輸出在形式與內容層面的可用價值，不涉及權威性、治理屬性或社會信任等宏觀議題，此類內容將在後續章節單獨展開探討。

2.1 AIGC 內容品質的界定與評估標準

2.1.1 內容品質的核心要義

AIGC 應用場景下內容品質不再局限于語句通順程度，而是由多個維度共同構成的綜合

屬性集合。首要維度是資訊精準度，涉及事實類資訊的文本必須在資料參數、時間節點、專有名詞及邏輯推演上與可核驗的客觀事實相符，尤其在新聞編撰、政策解讀、醫療科普及科技知識傳播等領域資訊偏差會直接削弱整個系統的實用價值與風險管控能力。任務適配性與語境契合度是另一重要維度，同一內容在不同任務場景中或許會獲得截然不同的評判，技術說明文檔需保證概念界定清晰、操作步驟完整，面向公眾的科普文本則更強調層次清晰、通俗易懂且規避過度專業化表述，生成內容能否匹配具體任務需求與目標受眾的理解水準本身就是品質評判的關鍵維度。

內容的結構編排與語言表達同樣是品質的核心構成，AIGC 生成結果若要真正被使用者接納通常需要在段落組織、論述節奏、句法規範及專業術語運用上達到基礎標準，結構鬆散或內容冗餘堆砌會導致讀者難以捕捉文本核心脈絡，即便資訊本身無誤也會降低實際可讀性與使用意願。創意呈現與非機械複刻也不可忽視，在廣告文案創作、敘事文本撰寫、課程內容設計等場景中生成內容過度依賴範本或固定句式將難以滿足用戶對差異化與新鮮感的需求，適度的表達變異、情境化描述及風格辨識度都是評估 AIGC 內容品質的重要依據。合規性與風險可控性是品質範疇的基礎要求，主要指向內容層面的底線準則如規避仇恨言論、歧視性表述、直白暴力內容或明顯侵權資訊等，責任劃分與監管機制等規範性議題將在後文從制度維度另行探討。

綜合上述維度可知 AIGC 內容品質本質上是多維度權衡的結果，不同應用場景會在精準度、可讀性、創造性與風險管控之間形成不同的組合側重。品質評估的首要工作是明確特定場景下哪些指標被列為優先考量項，哪些是可在合理範圍內調整讓步的次要指標。

2.1.2 品質評估的核心標準具體實施層面

AIGC 內容品質的評估通常採用量化指標與質性判斷雙路徑結合的方式。系統會借助一系列可量化的技術指標捕捉語言與結構層面的特徵，例如利用困惑度評估模型對後續詞彙的預測穩定性，借助 BLEU、ROUGE 等指標判定生成文本與參考文本的相似性及覆蓋範圍，或

通過詞彙豐富度與句式變化程度判斷文本是否存在過度範本化問題；對於包含事實陳述的內容部分系統還會設計與外部資料庫或知識庫的比對流程，將關鍵表述與已知權威資料進行匹配進而構建一套近似準確率與可驗證性的量化評估機制。

諸多與創造性、專業性及語境適配性相關的特徵仍需依託人工判斷來精準把握，編輯、審稿人員或專業評審會從結構明晰度、論證完備性、風格統一性及語用恰當性等角度對生成結果展開細緻審查；在醫療、法律、金融等高風險領域專業人員還會核查術語運用是否精準、推理鏈條是否存在斷裂、結論是否超出資料支撐範圍。實際操作中單一指標往往難以覆蓋全部需求，較為常見的做法是將自動化評估作為前置篩選工具再輔以抽樣式或重點式人工評估，兩者互補形成兼具效率與嚴謹性的品質評判體系。

2.2 AIGC 品質管控的主要技術與路徑

厘清合格內容界定標準後品質管控核心聚焦於通過系統設計與工作流程規劃讓生成結果穩定貼近既定標準。當前實踐中品質管控呈現多層次疊加格局，自動化檢測、人工審核、資料驅動調整與風險敏感領域強化管控相互協同，構成從前端生成到後端修正的完整管控鏈條^[5]。

模型輸出文本後系統即刻開展語法與拼寫校驗，標記明顯錯誤或不合語法規範的片段，同時借助語義相似度計算與文本匹配技術核查生成內容與訓練資料或外部語料在短語層面的重合程度，借此管控抄襲隱患與過度複刻問題。部分場景設定敏感詞庫與結構化準則，對涉及特定人群、議題或表述形式的內容實施預警與攔截。這一層級檢測主要針對形式特徵與顯性問題，優勢體現在成本低廉且回應迅捷。

人工審核與專業評估構成品質把關的二次屏障。普通內容由編輯或審核人員集中評估文本是否契合預設用途、語氣是否妥當、是否存在機器難以識別的語義模糊與價值偏差；高風險或高度專業化內容需引入醫師、律師、教師或領域研究者等專業群體，就生成結果中的專業表述與推理邏輯給出具體修改建議。這類人工介入雖無法覆蓋全部輸出，但可針對性聚焦

代表性樣本與關鍵節點，在資源有限前提下最大化提升整體品質與安全閾值。

越來越多系統依託使用者行為資料與主觀回饋優化生成策略，點擊率、停留時長、完讀率及退回刪除操作等均作為反映內容實際使用情況的間接指標。通過 A/B 測試對比不同風格、長度或結構的生成版本系統逐步鎖定特定場景下更易被接納的呈現形式；搭配簡易評分或標注機制讓使用者直接標記“有說明”“資訊失准”“表述晦澀”等問題類型，可在不顯著增加用戶負擔的前提下為後續模型微調與資料清理提供細緻指引。

面向醫療、金融、未成年人相關內容等敏感領域實踐中額外搭建風險強化型審查流程。這類流程通常對輸出內容施加更嚴格的結構與語用約束，例如限制模型出具具體診斷或投資建議轉而採用一般性資訊與風險提示的呈現形式，或明確涉及關鍵決策建議的文本必須經人工或專家覆核後方可對外發佈。同時針對這些領域單獨開展模型微調，使其表達更傾向於保留不確定性與條件限制，減少過度肯定式表述。從品質管控視角看這本質上是依據不同場景的風險承受能力設計匹配的內容生成與審查深度。

2.3 現有品質管控機制的技術瓶頸

技術實現與流程設計層面的多層次防護大幅增強 AIGC 內容的穩定度與實用價值，當前階段品質管控仍受限於若干無法規避的技術桎梏。最核心桎梏源自生成機制本身的概率屬性，大型語言模型依託條件概率分佈推演後續詞彙，輸入提示與參數配置完全一致依舊可能輸出差異明顯的結果。這種內在隨機性決定品質波動無法徹底消除，僅能通過反復參數調試與後期修正降低錯誤發生頻率卻難以將風險歸零。

當前自動化評估工具大多局限於淺層特徵識別，對句法失誤、基礎語義矛盾與明顯重複內容的辨識效果較佳，面對論證完備性、推理邏輯性、例證合適性等深層次內容判斷演算法至今仍難以輸出穩定可信的評估結果。這使得高階品質管控在較長週期內不得不持續依託人工審查，人工資源在規模化內容流通場景中必

然面臨成本與效率的雙重制約。

訓練資料的偏差對輸出內容品質形成深遠影響，即便管控環節著重強調資訊準確、內容多元與風險防範，模型訓練階段仍可能從失衡或帶有偏向性的語料中習得特定的敘事範式、視角取向或評價體系。這類偏向未必以顯性錯誤形態呈現，更多表現為風格固化、範式複製與視野局限等問題進而增加後期人工修正與二次編輯的工作量。

人工審核體系自身存在可持續運行與標準統一的難題，當內容生成效率遠超審核承載能力時即便實施抽樣審核與分級把關策略依舊難以杜絕部分問題內容的疏漏。不同審核人員在專業背景、價值取向與風格偏好上的差異會導致對同一文本做出截然不同的評判，進而影響系統長期優化的方向與穩定性。這些技術局限與人力約束共同構成當前 AIGC 內容品質管控無法回避的結構性瓶頸。

2.4 未來 AIGC 品質管控機制的優化路徑

立足上述技術與流程約束未來 AIGC 品質管控的發展走向更傾向於從「外掛式核驗流程」轉型為「深度嵌入模型架構與生產流程的全週期管控體系」。一個關鍵發展維度是在模型底層架構中植入更完善的事實核驗與知識約束模組，讓內容生成過程天然整合對外部權威知識庫的實時調取與資訊比對功能。包含專業知識的輸出內容生成過程中模型可同步調取結構化資料或權威資料庫核驗核心表述，視實際需求優化措辭表達或添加不確定性提示語，有效降低「文本流暢度達標但事實性存在偏差」的問題發生率。

另一核心升級路徑是把品質導向的各類回饋信號更體系化地融入模型訓練與參數微調全流程。綜合運用人類標注的偏好序列、任務達成度指標與長期使用者回饋資料，模型能夠在掌握語言生成規律的過程中逐步構建內部的「高品質內容判定體系」，進而在生成階段自主優化輸出風格與資訊密度配置。同期推進品質管控架構的分層化與協同化升級明確機器預審核、人工覆核與專家抽檢的職能邊界，讓各層級管控環節專注於適配自身的問題類型，實現成本可控前提下關鍵內容深度審查的持續落

地。

第三大核心方向是推動使用者回饋的系統化整合與高效應用使其在未來品質管控體系中佔據更關鍵的地位。通過搭建簡潔直觀的回饋入口持續歸集資訊錯誤、表達晦澀與風格失配等實際問題案例，系統可在反覆運算運算與參數更新中不斷校正模型固有偏向，推動品質管控從靜態的「標準條目集合」升級為與使用場景同步演進的動態適配機制。從長遠發展視角分析只有當品質管控邏輯真正深度融入模型架構設計、資料治理體系與全流程生產環節，AIGC 內容生產才能在保留創新活力的基礎上持續維持行業所需的穩定性與結果可預期性^[6]。

3 AIGC 生成內容的權威性與可信度分析

3.1 AIGC 權威性與可信度的概念界定

AIGC 深度融入知識生產與資訊流通場景中“權威性”與“可信度”不再局限於內容屬性範疇，而是演變為使用者在具體情境中完成的判斷過程。權威性核心指向知識是否被認可為“合規”“專業”“可信賴”，可信度更多體現為個人獲取資訊時形成的主觀確信度。兩者不存在簡單對等關係，部分內容可能在形式層面具備高度權威性卻未必獲得使用者的實質信賴，一些來源不明的生成內容也可能在具體場景下被認定為“足夠可信”^[7]。

AIGC 區別於傳統知識生產體系不依賴清晰的作者標識或機構擔保，其權威性難以通過“來源可查證”的方式予以確認。這一前提下使用者對內容的判斷不再以作者或機構為核心依據，更多依賴內容本身的呈現形態、語言範式與邏輯框架。AIGC 的權威性並非外部賦予而是在資訊接收過程中持續建構而成，這一建構過程並非單次完成而是融入個人的使用體驗、媒介認知能力以及已有知識體系之中。

可信度的生成同樣具備鮮明的主體性特質。使用者判斷 AIGC 內容可信度時通常不會開展嚴謹的事實核驗，而是依據內容的連貫性、表達的篤定程度以及與自身經驗的匹配度快速形成判斷。這一判斷模式並非“非理智”，而

是在資訊高度飽和環境中形成的實際適應策略。日常應用場景下可信度更多體現為“可接納性”，即內容是否能夠納入個人的理解框架進而用於指導行動或決策。

3.2 AIGC 權威性的主體內化路徑

AIGC 生成內容可在缺失明確知識溯源依據的前提下建立權威屬性，核心邏輯在於其權威屬性並非由外部賦予而是經由使用者的持續應用與心理內化過程逐步構建。使用者多次接觸生成內容過程中會逐步形成對其語言範式與接聽模式的熟悉度，這種熟悉度本身即構成信任的底層基礎。生成內容在各類場景中呈現“邏輯自洽”“回應適宜”特徵時其權威屬性便無需額外的外部佐證。

這一過程中主體並非被動接收資訊而是在持續優化自身的判斷維度與標準，原本需要依託權威來源進行核驗的內容逐步被替換為對“表述是否周全”“邏輯是否貫通”的直觀評估。長此以往權威的內涵便從“來源具備可信度”轉變為“回應契合預期設想”，這種認知轉變並非由平臺強制引導而是在個體層面潛移默化地發生，屬於典型的自我調適機制。

這種權威內化過程並不代表主體徹底喪失獨立判斷的能力，大量使用者在實踐過程中構建起一套“經驗導向的篩選體系”，例如針對涉及專業風險的問題保持審慎態度對日常資訊則給予相對較高的信任權重。這種差異化的判斷模式表明 AIGC 權威屬性的形成並非單向線性過程，而是與具體使用場景及個體經驗儲備存在緊密關聯。

3.3 可信度判斷中的自我治理邏輯

相較於權威屬性可信度更直觀地呈現為主體層面的自我治理行為，缺失明確外部規範約束的前提下使用者需要獨立承擔判斷內容真偽與潛在風險的責任。這種責任並未以制度化的形式予以明確而是以內在化的方式滲透到日常使用行為當中，個體在使用 AIGC 的過程中往往會持續進行自我提示如“不可全盤採信”“需自主再次判斷”，這類自我提示本身就是一種治理機制的具體表現。

實際應用過程中可信度的判斷並非通過系統性的核查流程完成，而是借助對內容“是否

具備合理性”的快速評估來實現。使用者並非追求內容的絕對正確性而是尋求一種具備可操作性的確定感，這種確定感並非源自外部權威的背書而是來自主體對自身判斷能力的內在信任。本質而言可信度的維繫依賴於主體持續確認“自身具備判斷能力”的主觀感知，而非內容本身所具備的客觀真實性。

這種自我治理邏輯在一定範圍內緩解了資訊不確定性引發的焦慮情緒，同時也可能掩蓋潛藏的各類風險。主體過度依賴自身經驗開展判斷時錯誤資訊並不會被及時識別，反而可能在多次使用過程中被逐漸合理化。由此可見可信度並非一種穩定不變的狀態，而是一個持續被協商、被調整修正的動態過程^[8]。

4 AIGC 內容的驗證與評價機制

4.1 AIGC 內容驗證的現有技術與實現路徑

AIGC 生成內容的驗證環節是保障其真實屬性與可信特質的核心支撐，人工智慧在內容創作與傳播領域的應用場景持續拓展，通過技術手段實現 AI 生成資訊的識別核驗與溯源已成為學術研究與產業實踐領域的共同聚焦方向。當前 AIGC 內容驗證主要依託多層級技術路徑與演算法架構，涵蓋文本一致性檢測、事實核查演算法、模型特徵識別及多源交叉驗證等實施模式。

文本一致性與語義匹配檢測屬於基礎層級的驗證手段，這類方法通過比對生成內容與權威資料庫或既有文獻資源的語義相似程度評估其是否與客觀現實事實保持一致。部分自然語言處理系統運用語義匹配演算法或上下文相似度計算（如余弦相似度演算法）對 AI 生成文本開展事實一致性校驗，該方式廣泛應用於新聞生成與科學文本審查場景能夠有效識別模型輸出中的“偽事實陳述”與邏輯矛盾，卻仍受限於資料覆蓋範圍與演算法識別精度。

事實核查演算法在 AIGC 驗證體系中佔據核心地位，這類技術主要依託知識圖譜、關係抽取與實體連結等手段將生成內容中的核心資訊要素（如時間座標、事件脈絡、人物資訊、數值參數）與權威知識庫進行匹配校驗。學術文本或政策分析生成場景中可通過知識圖譜比

對機制判斷文本引用資料的真實屬性，近年來部分 AI 實驗室已開發自動事實核查系統（如穀歌事實核查工具）能夠對生成文本中的關鍵陳述進行自動標注與證據檢索提升驗證流程效率，該技術仍存在語境依賴難題針對隱含性推理或複雜敘事場景的核驗能力相對不足。

模型特徵識別與生成軌跡溯源是區分 AI 生成內容與人工創作成果的重要手段，AIGC 生成內容往往具備獨特的語言分佈特徵、句式結構與概率模型可通過統計學習演算法識別其中的“模型特徵標識”。OpenAI 與斯坦福大學聯合研發的“AI 文本浮水印技術”通過在生成語料中植入特定統計特徵實現後期演算法檢測識別，這類方法為 AI 內容的可追溯性提供技術支撐卻同時面臨反向規避與模型模仿攻擊的實際挑戰。

多源交叉驗證機制在高可信度需求場景中的受關注程度持續提升，該方法通過整合多管道資訊來源（如資料庫資源、新聞機構素材與人工審核結論）對 AI 生成內容開展綜合驗證。科學研究與政策諮詢領域中多源交叉驗證常與人工評估相結合構建“人機協同”的驗證體系，確保內容既在資料層面準確無誤又在邏輯層面保持連貫自治。

人工參與及專家評估機制仍是 AIGC 驗證體系中不可或缺的構成部分，針對複雜模糊或高風險內容（如醫學診斷報告、學術論著、法律文書）單純依靠演算法驗證難以保障精準程度需引入人工評估流程結合專家專業知識開展事實核驗。專家審查在 AI 新聞編輯、科技報告撰寫與學術出版場景中應用廣泛，是提升 AI 生成內容權威屬性的關鍵補充手段^[9]。

4.2 AIGC 生成內容的評價體系與核心標準

AIGC 生成內容的評價體系與核心準則是衡量其品質等級、可信屬性與社會影響邊界的核心依據，區別於傳統人工創作模式 AIGC 內容生產依靠演算法邏輯與資料訓練構建，評價過程需同步考量技術性能表現、內容核心屬性與倫理規範要求三個維度。當前 AIGC 評價體系正逐步從單一技術評估轉向綜合性標準框架，形成以內容品質等級、事實精準度、可解釋屬性與社會責任履行度為核心的多層級評價架

構。

內容品質評價是 AIGC 評估的核心基礎維度，主要聚焦生成內容的語言表達流暢度、邏輯連貫度與語義合理性，常用自動化指標包括困惑度、BLEU 評分（用於文本生成精準度評估）及 ROUGE 指標（用於摘要撰寫與文本相似性判斷）。實際應用場景中研究機構通常結合機器評估與人工打分，通過設立“內容流暢度”“結構連貫性”“語義完整性”等評價維度構建綜合計分框架，新聞媒體的 AI 寫作系統會依據語義連貫度與讀者回饋資訊綜合判定內容品質等級進而優化模型輸出策略。

事實精準度與知識可靠性評價是保障 AIGC 內容可信屬性的核心支撐，該維度側重驗證生成資訊是否與客觀事實相符、是否存在虛假表述偏差內容或未經證實的資訊點。典型評價方法包括基於知識圖譜的事實比對機制、事實覆蓋比例與錯誤發生率指標，部分 AI 平臺近年引入“可信來源占比”指標用於評估生成內容中引用來源的權威屬性，這一方法已應用於科學論文自動撰寫與政策報告生成系統用以確保模型輸出契合學術規範與政策要求。

模型可解釋屬性與透明特質評價逐步成為 AIGC 標準體系的核心構成要素，AI 生成內容的可解釋程度直接影響用戶信任度與社會接納度需建立模型透明化評估機制。評價指標通常涵蓋生成依據可追溯性、演算法決策透明度與內容標識機制，部分 AI 平臺要求在生成內容中自動添加“AI 產出標識”明確其非人工創作來源以保障資訊傳遞的透明程度。

倫理與社會責任準則為 AIGC 評價提供規範支撐基礎，該維度強調生成內容是否契合社會價值導向與倫理底線要求、是否存在歧視性表述誤導性資訊或侵犯隱私的潛在風險。國際上較為成熟的框架包括歐盟提出的《可信 AI 倫理準則》突出公平屬性、問責機制與可審查特質；中國相關研究機構正探索 AI 內容倫理評估指標如“社會風險評估指數”“文化適配度評分”等用以衡量生成內容的社會接納程度。

綜合評價框架的形成標誌著 AIGC 標準體系的系統化發展趨勢，當前多個國家和機構

已提出多維度綜合評價模型例如 IEEE 的《AI 系統品質評估標準》（P7000 系列）與中國信通院的《人工智慧內容品質評估指南（2024）》等。這些標準普遍採用“技術層面—內容層面—倫理層面”三維度架構強調品質與責任兼顧的評價原則。

4.3 驗證機制的現實難題與創新趨向

AIGC 內容驗證機制的發展雖已實現顯著進展但技術落地、標準規範與社會治理等維度仍存在多重困境。這些困境既會影響 AI 生成內容的可信屬性也決定人工智慧能否真正融入科學知識生產與公共傳播體系，技術與治理機制的創新探索持續推進驅動驗證機制向體系化、多層級與智慧協同方向穩步發展。

事實驗證的複雜屬性仍較為突出，AIGC 生成內容常整合多領域知識邏輯關聯緊密且語義層次豐富，傳統事實核查演算法難以精準辨識語境關聯的虛假資訊。模型應答科學或醫學類問題時可能生成表面合理但事實存在偏差的內容，這類“偽事實表述”比顯性虛假資訊更難甄別。當前驗證模式主要依託知識庫匹配與語義比對，生成內容涉及隱含推理或跨學科知識時其正確性難以通過自動演算法全面核實。

多模態內容驗證尚未建立統一規範，AIGC 在圖像、語音、視頻等領域的應用持續拓展內容驗證的複雜程度不斷攀升。文本驗證依託語義比對邏輯，圖像與音訊內容則需通過特徵識別、中繼資料追蹤等方式實施驗證，目前尚未構建跨模態統一驗證框架，文字說明與配圖間的事實偏差往往無法通過單一演算法檢測致使多模態生成內容的真實屬性驗證仍處於起步階段。

演算法的不透明屬性削弱驗證體系效能，多數 AIGC 系統屬於商業閉源架構外部機構難以知曉其訓練語料、參數配置及生成邏輯，致使驗證機構無法追溯內容生成路徑。生成內容出現偏差時即便發現問題也難以明確責任主體或定位錯誤環節，這種“黑箱化”困境在學術內容與政策解讀等高風險領域表現尤為顯著加大內容監管的制度實施難度。

倫理與法律規範的缺失加劇驗證困境，AI

生成內容涉及誤導性資訊、隱私洩露或價值偏向時現行監管體系缺乏明確法律約束。部分平臺以“技術創新”為藉口規避內容審核致使 AI 生成內容在社會傳播中難以追究責任；驗證機制在倫理判斷層面存在“機器無法界定價值取向”的困境，演算法可判別資訊真偽卻難以判斷內容是否“適宜生成傳播”。

面對上述現實困境 AIGC 內容驗證機制正呈現多重創新方向。新一代生成模型逐步嵌入“事實比對”與“內部回饋迴圈”功能，在生成過程中自動調取外部資料庫實施即時驗證，模型發現輸出與既有知識不一致時可實現自我修正從源頭減少錯誤資訊傳播。基於區塊鏈技術的內容溯源系統持續完善，通過為 AI 生成內容分配唯一數位簽章與時序標識構建完整的“生成—傳播—驗證”鏈路，確保每段內容均可被追蹤與審計，該機制在學術出版、政策檔與媒體新聞領域適用性尤為顯著。人機協同驗證機制逐步落地，高風險或高專業度領域中引入人工專家與演算法系統協同審查，AI 負責初步篩查與一致性檢測專家承擔語義與事實覆核工作，實現驗證效率與精確度的動態平衡。社會化驗證模式持續探索，部分科研機構與媒體平臺嘗試搭建“開放驗證平臺”允許公眾、研究者與協力廠商機構參與 AI 內容評估與監督，形成動態回饋機制提升社會信任水準。

4.4 未來驗證與評價機制的發展展望

AIGC 技術持續快速反覆運算其生成內容的驗證與評價機制正經歷從“事後核驗”到“全流程管控”的轉型歷程。未來驗證與評價體系不再局限於輸出結果的核驗工作而是延伸至演算法架構設計、資料來源頭篩選、生成邏輯推演及社會影響評估等多個維度，實現動態化、透明化與多維度的品質管控訴求。

驗證機制將呈現“模組內嵌”與“流程前置”的發展態勢，以往驗證多依託人工或獨立系統實施事後檢測，未來 AIGC 系統將依託模型內嵌的驗證模組在生成流程中即時監測事實匹配度、來源可信度與邏輯連貫性，實現“生成即驗證”的核心訴求。這一發展方向已在部分科研類 AIGC 平臺落地實踐，比如通過語義匹配與知識圖譜比對即時識別潛在偏差在

生成階段降低虛假資訊傳播概率。

驗證機制將與評價體系深度融合構建“技術—倫理—社會”三位一體的綜合標準體系。未來 AIGC 評價體系既關注生成內容的語言水準與事實精確度也會評估其社會職責、倫理風險與文化適配性。科學知識生成領域中 AI 產出的論文或技術報告需通過多維指標體系，涵蓋學術引用有效性、資料可追溯性、演算法透明度與倫理合規性等保障契合科研規範與公共信任訴求。

評價機制將更多依託開放協同的治理架構，政府監管部門、科研機構、企業平臺與公眾使用者將共同參與 AIGC 內容的監督與回饋工作搭建“多主體共治”的社會化評價網路。這一機制可借助開放資料庫、可追蹤報告及跨平臺審計實現持續動態監管，降低 AI 濫用與內容失真風險。

技術創新將成為驗證與評價機制演進的核心動力支撐，區塊鏈溯源技術、聯邦學習演算法與可解釋性模型等將助力構建去中心化驗證體系實現內容來源可查、修改軌跡可溯、相關責任可究。未來 AIGC 系統有望具備“自我驗證”與“自我評估”功能通過內部演算法自檢持續優化生成品質，依據外部回饋動態調整模型參數。

總體來看 AIGC 驗證與評價機制的未來發展方向將是標準體系化、技術智慧化與治理協同化的三位一體。唯有構建從演算法層面到社會層面的全鏈路可信體系才能保障人工智慧在知識生成與公共傳播領域既具備創新活力又擁有權威屬性與可控特質^[10]。

5 AIGC 生成內容的評價體系與標準

5.1 AIGC 內容的評價標準與核心框架

AIGC 生成內容的評價標準與核心框架是保障人工智慧在科學知識生產與社會傳播中具備權威特質與可信屬性的關鍵支撐，不同於傳統人工創作模式 AIGC 的生成邏輯由演算法主導，其內容品質既依賴語言表達效果與邏輯連貫程度更取決於演算法架構設計、資料來源頭選擇與生成流程的透明屬性，建立科學系統可衡量的評價標準對保障 AI 生成內容的真實

特質、可追溯屬性與社會責任具有重要意義，當前 AIGC 內容的評價標準主要包含四個核心維度即事實精準度、邏輯連貫特質、倫理合規屬性與公共可信程度，這些指標共同構成 AIGC 內容品質的多層級評價架構。

事實精準度是 AIGC 內容評價的首要基準，AI 生成系統通常依託大規模訓練語料開展知識提取與語言生成工作，模型存在事實偏差、資訊滯後或語義謬誤的潛在風險，事實精準度要求 AI 輸出內容與現實世界資訊保持契合，可通過多源驗證機制實現例如與權威知識圖譜比對、事實資料庫交叉核驗或參考文獻檢索等方式，科研領域中 AI 生成的論文摘要應準確反映原始研究的資料結論與研究方法，不得出現偽造引用或編造資料的情況，事實核驗機制的引入既是技術層面的保障更是 AI 生成內容獲得社會信任的根本前提。

邏輯連貫特質是 AIGC 內容品質的內在基準，聚焦生成內容內部結構是否合理、推理過程是否自治、論述表達是否連貫，邏輯謬誤常出現在多步推理、知識遷移或概念整合過程中，尤其 AIGC 生成科學性內容時邏輯連貫特質直接影響知識的科學屬性與可解釋程度，可採用語義鏈分析、推理路徑視覺化等方法檢測模型生成過程中的邏輯連貫程度，進而識別潛在的推理漏洞或資料雜訊引發的邏輯斷裂。

倫理合規屬性是 AIGC 評價體系中的社會規範維度，AI 生成內容時必須遵守法律法規與社會倫理準則，避免傳播虛假資訊、歧視性表述、偏見觀點或侵犯隱私的內容，倫理評價應覆蓋三個層面即演算法偏見檢測、生成內容核查與責任歸屬界定，例如 AI 生成的科學報導應避免強化性別偏見、科研國家主義或學科歧視等隱性傾向，實踐過程中可通過引入倫理約束演算法、價值敏感設計（VSD）原則與人工覆核機制，保障 AIGC 在知識生產中的社會責任與道德合法性。

公共可信程度是評價 AIGC 生成內容能否被公眾與學術界廣泛接納的重要指標，可信程度既源於事實與邏輯的精確屬性也取決於內容來源的透明程度與模型輸出的可解釋性，AI 系統能夠明確標注資料來源、模型版本與訓練範圍時使用者對內容的信任度會顯著提升，未

來 AIGC 的可信程度評估應結合用戶行為分析、社會回饋機制與協力廠商審查制度，通過持續性評價構建動態信任體系。

整體來看 AIGC 生成內容的評價框架可分為三個層面，第一層為技術層面即模型內部的自動化自檢機制對生成內容開展即時監測與修正，第二層為平臺層面包含企業或科研機構的多重審查機制負責技術核查、倫理監督與內容覆核，第三層為社會層面即由公眾、學術團體與監管機構共同參與的外部評估體系，保障 AI 內容符合公共利益與科學規範。

這一三層結構的框架既實現了從“演算法生成”到“社會認定”的完整閉環也為 AIGC 在科學知識生產中的標準化與制度化提供了基礎路徑，未來隨著國際 AI 治理體系的發展中國應結合自身科研與文化語境，建立具有本土特色的 AIGC 內容評價體系實現科學屬性、規範特質與社會價值的統一。

5.2 評價指標的設計與落地實施

AIGC 生成內容的評價指標設計與落地實施是將抽象品質標準轉化為可落地可衡量評估架構的核心支撐，AIGC 生成內容在科學知識傳播、媒體資訊發佈與教育場景應用等不同領域呈現多樣化特徵，其評價指標體系需兼顧通用屬性、層級特質與動態適配能力，整體劃分為技術層指標、內容層指標與社會層指標三大類別構建層級遞進的綜合評估結構。

技術層評價聚焦模型性能表現與生成流程的可控屬性，核心指標包含①事實契合率測算生成內容與權威資料及知識庫的契合水準；②邏輯連貫水準借助語言模型的語義相關度評分與推理鏈解析判斷內容邏輯自治屬性；③可解釋性分值評估模型生成過程中對資訊源頭、推理依據的披露程度；④可追溯性係數通過記錄資料來源與生成路徑保障結果可複查特質，這些指標設計使 AI 系統輸出內容時實現“自我驗證”有效降低錯誤傳播風險。

內容層評價指標圍繞 AIGC 生成文本或圖像的品質核心展開，涵蓋①資訊精準程度依託事實核驗與資料比對作為核心方式；②語言合規程度考察生成內容在語法規範、結構架構、語義表達層面的精準屬性與流暢特質；③原創

屬性與創新特質通過相似度檢測與語義重構解析研判內容是否具備獨立表達與知識增值效果；④倫理合規屬性識別內容中是否存在歧視傾向、偏見表述、隱私侵犯或倫理隱患，科學知識生成領域中原創屬性與倫理屬性是保障內容學術合法特質的重要指標。

社會層評價指標聚焦公眾信任程度與社會影響範圍，包括①可信程度分值結合用戶信任回饋、專家覆核建議與社會平臺聲譽研判 AI 生成內容的整體可靠水準；②社會影響係數通過輿情監測與資訊擴散模型研判內容傳播的正負向效應；③透明程度指數測算 AIGC 在資訊來源、演算法邏輯與責任歸屬方面的公開程度，這類指標落地體現“技術—倫理—社會”一體化治理思路實現 AI 品質管控與社會責任的同步推進。

落地實施階段 AIGC 評價指標採用多主體協同與分階段推進的機制，系統內部自動化評估先行 AI 模型通過演算法自檢輸出初步品質分值，平臺與專家評估系統後續覆核針對科研資料、醫療建議等高風險內容開展人工審查，社會化回饋機制同步引入結合用戶舉報、媒體核查與監管機構評估形成動態更新，各階段回饋資料反向訓練模型持續優化生成策略與判斷邏輯。

指標權重設置需根據不同應用場景靈活適配，科研文本應強化“事實契合率”與“原創屬性”的權重占比，面向公眾的資訊傳播內容需提高“倫理合規屬性”與“社會可信程度”的比重分配，此外指標體系需具備動態適配特質隨技術演進與社會規範更新持續優化，未來結合大資料監測與人工智慧評估將構建自動化 AIGC 內容評分系統，實現從靜態核查到動態品質管控的轉型為 AIGC 在科學研究與社會應用領域的可信落地提供堅實支撐。

5.3 評價體系的優化路徑與技術難題

AIGC 技術持續演進現有評價體系面臨多重挑戰尤其體現在內容品質與可信程度的評估維度，提升 AIGC 生成內容的精准特質、可靠屬性與社會適配能力評價體系的優化是核心關鍵環節，優化過程並非簡單指標調整而是涉及技術難題、倫理研判與實施層面的複雜問題。

自動化評估的精准程度是評價體系優化中尤為顯著的技術難題，AIGC 生成內容具備較強動態特質與多樣屬性其品質評估難以單純依託固定規則或單一模型，現有自然語言處理（NLP）與電腦視覺（CV）技術在精准屬性與流暢特質上已有明顯突破卻仍難以全面應對語境解讀、隱含語義及文化背景差異等複雜情形，AIGC 生成內容可能語法層面無明顯疏漏但其中隱性偏見、誤導性表述或倫理失當傾向仍難以通過單一技術指標甄別，評價體系優化過程中需結合多種技術手段開展交叉驗證提升 AI 模型在複雜場景中的判斷能力這是當前需解決的核心問題。

模型可解釋屬性與透明特質是評價體系技術優化的核心方向，AIGC 生成流程的黑箱特質使其產出內容缺乏透明推理脈絡使用者難以明晰 AI 得出結論的具體路徑，這既影響內容的權威特質與可信屬性也給後期品質管控帶來極大困難，優化過程中如何讓模型生成內容時提供可解釋的推理流程如何搭建基於模型透明特質的評價標準已成為當前技術研究的熱點，這要求設計評價體系時採用更先進的演算法與技術推動 AIGC 模型向更“可解釋”方向發展進而增強其輸出內容的透明程度與信任水準。

倫理問題的多維度研判亦是技術優化中的重要難題，AIGC 生成內容的倫理合規屬性在醫學、教育、法律等敏感領域中尤為關鍵，現有評價體系檢測內容倫理問題時過於依賴人工標注造成效率偏低且易出現偏差，有效解決這一問題需在評價體系中引入更多倫理評估標準通過深度學習與語義解析等技術提升模型對倫理合規問題的敏感程度，AI 可通過自動化識別內容中潛在偏見、歧視性表述加以規避但這要求評價體系能對不同文化背景與社會環境中的倫理標準進行靈活適配。

跨領域協同與資料共用在優化評價體系過程中面臨一定技術阻礙，AIGC 應用場景持續拓展使得內容評價標準與技術手段需跨越傳統領域界限開展跨行業、跨學科的整合，不同領域對 AIGC 生成內容的需求與標準差異明顯保持行業標準獨特性的同時搭建統一評價體系與評估框架依賴多方技術協作與資料共用，此

過程中如何保障資料隱私、安全屬性及模型適配能力將成為亟待解決的技術難題。

實際實施層面評價體系的實操屬性亦是需優化的重要維度，部分現有評價標準實操性偏弱難以在實際應用中精準落地，生成內容評分系統中如何對長文本、圖片或音訊等不同形式內容進行統一合理的評分仍是技術性難題，不同評價標準對結果的評分權重分配不均且難以量化進一步加劇評價體系的複雜程度，未來優化工作應從簡化操作流程、提升自動化水準與豐富評估方法等角度出發確保評價體系的易用屬性與實施效果。

5.4 國際與國內 AIGC 評價標準的比較分析

全球範圍內 AIGC 生成內容的評價體系仍處於構建階段不同國家及地區在技術基準、倫理準則與治理架構上表現出明顯不同，整體來看歐美國家的 AIGC 評價體系更側重透明特質、可解釋屬性與責任可追溯屬性中國的評價標準則更強調內容安全屬性、倫理合規要求與社會影響管控，這種差異體現各國在技術治理理念、社會文化背景及政策導向層面的不同取向。

國際層面歐盟與美國在 AIGC 評價標準制定上處於領先地位，歐盟於 2023 年正式頒佈《人工智慧法案》(AI Act) 首次依據風險等級對 AI 系統實施分級管控，其中對 AIGC 類生成模型提出明確的透明特質與可追溯屬性要求，該法案明確生成式 AI 需標注內容源頭、說明是否為 AI 生成保障模型訓練資料的合法屬性與可審計特質，歐洲標準化委員會 (CEN) 與國際電子電機委員會 (IEC) 聯合發佈的 AI 內容評估框架強調技術中立屬性與資料可解釋特質，將事實契合程度、偏差甄別能力和倫理自我核驗列為核心評估維度，美國以市場機制為核心導向由產業聯盟與大型科技企業牽頭標準構建，美國國家標準與技術研究院 (NIST) 發佈的《人工智慧風險管理框架》(AI RMF 1.0) 提出從模型開發、內容生成到結果驗證的全過程評估原則，重點關注演算法偏差管控、內容真實屬性評估與責任歸屬界定機制，OpenAI、Google DeepMind 等機構已搭建內部 AIGC 品質管控體系通過多層人工審查、事實交叉核驗及偏差檢測工具保障生成內容的可靠屬性與安

全特質。

中國的 AIGC 內容評價體系更具政策導向與社會治理特徵，2023 年發佈的《生成式人工智慧服務管理暫行辦法》明確 AIGC 服務提供者須對生成內容承擔責任，確保輸出契合社會主義核心價值觀與公共秩序要求禁止生成虛假資訊、有害內容或侵犯隱私的內容，評價體系建設方面中國標準化研究院正在推進《人工智慧生成內容安全標準體系框架》的制定，核心涵蓋事實精準程度、內容安全屬性、模型可控特質和倫理合規要求四大指標，中國在 AIGC 監管中引入事前報備、事中監控、事後問責的分層管理機制與西方的行業自我約束 + 外部監管模式形成區別，百度、阿裡巴巴、智譜 AI 等企業已搭建內部多維內容審查系統結合語義識別技術、人工覆核流程與黑名單機制，實現 AIGC 輸出的自動篩查與風險分級管控。

國際與國內標準的核心差異體現在三個維度，治理邏輯存在差異歐美強調透明屬性與問責機制注重技術過程與結果的可解釋特質中國強調可控特質與合規要求更注重生成內容的社會影響與政治倫理安全，評估重點有所不同歐美的 AIGC 評價標準偏向技術導向關注演算法偏差、模型透明程度與資料來源合法屬性中國的評價體系以社會導向為核心重視內容是否符合法律法規、倫理道德與社會穩定要求，實施機制存在區別西方主要依靠行業自我約束與社會監督力量中國採用政府監管與企業自主核查相融合的管理模式形成更具集中性的管控體系。

儘管存在差異國際與國內在 AIGC 評價體系建設上呈現趨同態勢，均逐步重視生成內容的真實屬性與可追溯特質均在探索搭建跨學科倫理評估框架均強調模型透明程度與可解釋屬性的重要價值，未來中國可在借鑒國際經驗的基礎上搭建具有中國特色的 AIGC 評價架構，在保障內容安全與倫理合法的前提下加強對技術透明程度、資料合規要求與社會信任構建的平衡推動 AIGC 品質治理標準的國際化與科學化演進。

6 結論與展望

6.1 AIGC 內容的權威特質與品質管控現狀總結

近年 AIGC(人工智慧生成內容)技術持續高速發展帶動知識生產與資訊傳播模式發生深度變革,其在科研寫作、新聞採編、教育資源創制及輿情研判等領域展現出顯著應用潛力,伴隨技術應用範圍拓展 AIGC 內容的權威特質與品質管控問題逐漸成為影響其社會信任水準與學術價值的核心議題,當前 AIGC 內容品質管控體系尚未完全成熟在技術、制度與倫理維度均面臨多重挑戰。

技術維度上 AIGC 生成內容的精准屬性與連貫特質仍有明顯差距,模型訓練依託大規模互聯網資料資訊源頭存在真實程度不一、更新遲緩與偏見積澱等情況,導致 AI 輸出內容可能出現事實謬誤、邏輯斷層與語義偏差,尤其科研知識生成與政策分析等高精度領域 AI 生成內容若缺乏嚴格驗證機制極易造成偽知識或誤導性結論擴散,近年雖引入知識圖譜對應、事實核驗演算法與模型溯源技術但當前自動化品質檢測仍以語義流暢度為主,難以全面評估內容真實屬性與專業深度。

制度與治理維度 AIGC 內容的監管架構正在逐步構建但仍呈零散狀態,國際層面主要採用“行業自我約束 + 政府監督管理”的融合模式,中國則在政策層面搭建了更為嚴格的內容審查與報備機制,如《生成式人工智慧服務管理暫行辦法》明確了 AI 內容的安全邊界與主體責任,這些管理舉措在執行過程中仍面臨標準不統一、評價指標缺乏實操屬性等問題,不同平臺、行業間的品質標準存在明顯差異導致 AIGC 的可信程度難以實現跨平臺統一。

倫理與社會維度 AIGC 的權威特質受到演算法偏見與責任界定模糊的雙重作用,AIGC 生成內容常被視作“客觀產出”實則體現訓練資料與演算法架構中蘊含的人類偏見與價值傾向,知識傳播領域中這種“偽中立屬性”可能掩蓋不平等敘事或強化既有權力結構,削弱內容的科學公正特質同時 AIGC 的責任歸屬問題尚未明確,當生成內容引發倫理爭議或造成社會誤導時責任主體的界定仍存在模糊空間。

當前 AIGC 內容品質管控已呈現積極發展態勢,大模型企業與科研院所正在推行多層級驗證機制包括事實比對、專家覆核與社會回饋系統以提升內容可信程度,部分平臺已開始探索自動化倫理審查與語義安全評估架構通過演算法識別潛在風險並在生成階段進行即時修正,越來越多跨學科研究正嘗試構建“科學—倫理—社會”三維評價架構在技術進步與公共利益之間尋求平衡。

6.2 政策建議與行業發展導向

應對 AIGC 內容權威特質與品質管控相關挑戰中國及全球的政策制定與行業演進需從技術治理、制度構建與社會協同三個維度系統推進,AIGC 作為知識生產的核心工具其健康演進既依賴演算法與模型的優化更取決於法律法規的健全、行業標準的規範落地以及倫理與社會監督的有效參與。

政策維度應加快搭建國家級 AIGC 內容品質管制與評估架構,制定統一的《AIGC 內容品質與安全基準》明確事實核驗、資料合規要求、倫理核查與可解釋屬性等基本準則形成跨行業跨平臺的品質管控標準體系,推動“生成式人工智慧透明化條例”落地實施要求 AIGC 產品明確標注內容生成源頭、演算法版本及資料訓練範圍保障公眾知情權與可追溯特質,政府建立 AI 生成內容報備制度與演算法安全評估機制確保 AIGC 模型上線前通過倫理核查與安全測試從源頭防範錯誤資訊與偏見傳播。

行業發展維度企業與科研院所需主動肩負 AIGC 品質治理的主體責任,建立企業內部“內容核查與倫理治理委員會”以專家參與模式對 AI 輸出內容開展定期抽查與風險評估,行業聯盟共同推動“AI 內容自主約束公約”在競爭中形成合規共識與共用責任,推動 AIGC 平臺間的資料共用與模型互認機制促進標準的互聯互通,科研院所強化 AIGC 模型的學術核查制度確保 AI 生成科研文本、技術報告等內容的可驗證屬性與可複現特質防範“AI 學術造假”現象蔓延。

社會治理維度需搭建多元主體參與的協同監督機制,政府監管部門通過智慧化手段(如

語義識別與演算法審計)實現對 AIGC 內容的動態監管,公眾與媒體納入社會監督體系通過舉報、回饋與輿論評估參與 AIGC 治理,積極推動公眾 AI 認知素養培育使社會具備識別與評估 AI 內容的能力形成“技術透明—公眾參與—責任共擔”的良性治理生態。

未來發展導向中 AIGC 行業應以可信 AI 為核心發展導向,加快推進可解釋人工智慧(XAI)技術研發使模型生成知識的同時提供邏輯依據與資料來源頭提升內容的可驗證特質,推動 AI 與區塊鏈技術融合構建基於溯源機制的內容認證體系實現從資料登錄到結果輸出的全過程追蹤,引導 AIGC 應用向高價值領域延伸(如科學研究、教育創新、公共服務等)在規範化基礎上實現社會效益最大化。

AIGC 的政策與行業發展需在“技術創新”與“社會信任”之間尋求平衡通過法制化、標準化與智慧化的協同路徑搭建開放安全負責任的 AI 內容生態系統,使 AIGC 真正成為推動知識增長與社會進步的核心力量。

6.3 未來研究的方向與潛在可能

AIGC 內容的權威特質與品質管控是動態發展的研究範疇,其發展取向既取決於人工智慧技術的突破也依託於社會治理理念、倫理規範與跨學科研究的推進,未來研究將聚焦多方面實現 AIGC 從“智慧生成”到“可信知識創造”的轉型。

搭建可解釋屬性與透明特質的知識生成機制是核心研究議題,目前多數 AIGC 模型生成內容時缺少明確推理脈絡與溯源記載,導致知識源頭不可追蹤、邏輯過程不可驗證,未來研究應聚焦“生成可解釋性”通過語義鏈視覺化、因果關係建模與資料來源映射,使 AI 生成的知識具備可追蹤、可驗證、可審計的特徵增強內容的科學屬性與可信程度。

倫理與價值融入機制需深入探析,AIGC 在跨文化跨領域應用中遭遇複雜價值判斷情形,模型訓練階段融入倫理約束與文化多樣認知直接決定其輸出的社會適配能力,未來研究

可從演算法倫理學、社會技術系統(STS)和人機協同治理等角度探索 AI 系統生成知識時的價值平衡機制,規避技術理性對社會多元價值的消解。

多模態內容品質管控與跨領域標準化評估架構是 AIGC 治理的關鍵取向,當前研究多聚焦文本生成範疇 AIGC 在圖像、視頻、音訊及科學類比資料等領域應用持續拓展,未來需建立統一品質評估架構涵蓋語義精準度、資料真實屬性、感知一致特質和倫理安全等維度推動多模態 AIGC 內容的協同治理,不同行業應聯合搭建跨領域品質基準與審查架構形成可比屬性與互認機制。

人機協同創作與智慧核查的聯動機制是未來研究新趨向,AIGC 持續參與科研寫作、政策起草與新聞編輯等高風險範疇單純依託演算法自動生成與核查難以滿足可信程度要求,未來研究應探索“人機混合決策模型”通過人工專家與 AI 系統的聯動學習實現對生成內容的雙重核驗與動態修正,在保障效率的同時提升內容品質與責任歸屬的明確程度。

AIGC 社會影響的長期評估與政策回饋機制是亟需關注的新範疇,AIGC 既是技術現象也是社會變遷的驅動力量未來研究應跨越技術邊界,關注其在公共輿論形成、知識權力結構、勞動市場與文化生態中的深層影響,通過搭建長期追蹤與政策回饋架構研究者可為 AIGC 的社會治理提供持續的理論支撐與實證基礎。

整體來看未來 AIGC 研究方向將從單一的“內容生成與品質核查”轉向綜合性的“知識治理與社會信任構建”,這不僅是技術優化問題更是融合人工智慧倫理、資訊科學、社會學與哲學的系統性研究議題,隨著 AI 能力的持續演進 AIGC 研究將逐步從“工具視角”邁向“認知共同體視角”,推動人類社會進入更具智慧特質、透明屬性與可信特質的知識生產新時代。

參考文獻：

- [1] 人工智慧生成內容(AIGC)白皮書編寫組. 人工智慧生成內容(AIGC)白皮書[M]. 北京:中國資訊

通信研究院, 2022.

[2] Cao Yihan, Li Siyu, Liu Yixin, Yan Zhiling, Dai Yutong, Yu Philip S., Sun Lichao. A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT[EB/OL]. arXiv, 2023, arXiv:2303.04226. <https://arxiv.org/abs/2303.04226>.

[3] Wang Yuntao, Pan Yanghe, Yan Miao, Su Zhou, Luan Tom H. A survey on ChatGPT: AI-Generated contents, challenges, and solutions[EB/OL]. arXiv, 2023, arXiv:2305.18339. <https://arxiv.org/abs/2305.18339>.

[4] 證券研究報告編寫組. 人工智慧系列深度報告：AIGC 行業綜述篇——開啟 AI 新篇章[R/OL]. 證券研究報告, 2023-03-20.

[5] 錢亮帆, 劉立翔. 高校 AIGC 檢測標準的反思：人工智慧與人的主體性讓渡[J]. 中國現代教育學報, 2025, 1(1): 25-31. DOI: 10.70693/jyxb.v1i1.9.

[6] Papagiannidis E, Alamanos E. Responsible artificial intelligence governance: a review and research agenda[J]. Technological Forecasting and Social Change, 2025, 196: 122727. DOI: 10.1016/j.techfore.2024.122727.

[7] Lu Y. Reforming copyright law for AI-generated content[J]. TechReg, 2025, (2): 39-58.

[8] Huschens M, Briesch M, Sobania D, Rothlauf F. Do you trust ChatGPT? Perceived credibility of human and AI-generated content[EB/OL]. arXiv, 2023, arXiv:2309.02524.

[9] Elkhatat A. M., Elsaid K., Almeer S. Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text[J]. International Journal for Educational Integrity, 2023, 19: 17. DOI: 10.1007/s40979-023-00140-5.

[10] Taeihagh A. Governance of Generative AI[J]. Policy and Society, 2025, 44(1): 1-23.

版權聲明

© 2025 作者版權所有。本文依據“知識共用署名 4.0 國際授權合約”（CC BY 4.0）以開放獲取方式發佈。該許可允許使用者在任何媒介中自由使用、複製、傳播與改編文章（含商業用途），惟須明確署名原作者及出處，並注明所作修改（如有）。完整協議詳見：<https://creativecommons.org/licenses/by/4.0/deed.zh-hans>

出版聲明

所有出版物中的陳述、觀點及資料僅代表作者及供稿者個人立場，與 Brilliance Publishing Limited 及/或編輯人員無關。Brilliance Publishing Limited 及/或編輯人員對因內容所提及的任何理念、方法、說明或產品所導致的人身或財產損害概不負責。

DOI: 10.53104/xdkxsts.2025.01.02.003

分層土壤中降雨驅動下重金屬遷移行為的層間擴散機制探討

趙天宇¹

1. 山西農業大學，山西 太原，030031

摘要：在長期重金屬污染條件下，分層土壤在垂直方向上表現出明顯差異，不同土層中重金屬逐漸累積。降雨作為主要的外部因素，會改變土壤孔隙中的水流狀態，使溶解態和絡合態重金屬在不同土層之間發生移動。本研究梳理了分層土壤結構和降雨入滲過程對重金屬遷移的影響，說明瞭對流移動、擴散作用、吸附與解吸過程以及指狀流動等多種過程之間的共同作用。通過理論模型分析發現，在降雨後期，濃度差引起的擴散仍然持續推動重金屬向深層土壤移動，並且在不同孔隙結構和降雨條件下，遷移速度和累積深度存在明顯差異。

關鍵字：分層土壤；重金屬遷移；降雨入滲；層間擴散；對流—彌散—吸附模型；優先流

Investigation of the Interlayer Diffusion Mechanism of Heavy Metal Migration Under Rainfall in Layered Soils

ZHAO Tian-yu¹

1. Shanxi Agricultural University, Shanxi 030031, P.R.China

Correspondence to: ZHAO Tian-yu; Email: tyzhaotianyu@outlook.com

Abstract: Layered soils under long-term heavy metal contamination exhibit significant vertical heterogeneity and multi-layer accumulation characteristics. Rainfall, as a primary external driving force, alters the pore water dynamics and promotes the redistribution and migration of dissolved and complexed heavy metals between soil layers. This study systematically reviews the effects of layered soil structures and rainfall infiltration processes on the migration behavior of heavy metals and elucidates the coupled mechanisms involving advection, dispersion, sorption-desorption equilibrium, and finger flow phenomena. The theoretical model analysis demonstrates that diffusion driven by concentration gradients continuously promotes the downward migration of contaminants during the post-rainfall stage and highlights the differences in migration rates and accumulation depths under varying pore structures and rainfall conditions. The results provide a theoretical basis for risk identification, zonal monitoring, and barrier control strategies in contaminated soils and have important implications for agricultural ecological security and groundwater protection.

Key words: layered soil; heavy metal migration; rainfall infiltration; interlayer diffusion; advection-dispersion-sorption model; preferential flow

收稿日期：2025-07-08 返修日期：2025-12-14 錄用日期：2025-12-23 出版日期：2026-01-06

通信作者：tyzhaotianyu@outlook.com

引用格式：趙天宇. 分層土壤中降雨驅動下重金屬遷移行為的層間擴散機制探討[J]. 現代科學探索, 2025, 1(2): 30-36.

引言

近年來，土壤重金屬污染問題愈發突出，工業排放、礦業開發和農業施肥等多重壓力持續向表層土壤輸入鎘、鉛、鋅等污染物。這些重金屬在土壤剖面中的累積與再分配，成為威脅地下水安全和農業生態系統的重要隱患。部分地區由於長期高強度工業排放和無序堆放固體廢棄物，表層土壤的重金屬含量已顯著超出農用地環境標準，污染向下擴散趨勢日益明顯^[1]。與大氣沉降和人為直接擾動不同，降雨過程作為一種持續且週期性的自然動力，對污染物垂直遷移具有決定性影響。

降雨入滲會明顯改變土壤中的含水情況和水流條件。溶解態重金屬會隨著水分流動，在不同深度的土層之間不斷移動。降雨的強度、持續時間和發生次數，會影響水分流量和重金屬隨水移動的速度。幹濕反復變化，會影響重金屬在土壤固相和水相之間的釋放和再次吸附。隨著水分逐漸向下滲透，表層中濃度較高的污染物會在濃度差的作用下發生擴散，並逐步在下部土層中累積，或通過優先流通道直接進入更深層土壤。有研究指出，強降雨事件可能在短時間內導致深層土壤中污染物快速增加，並形成難以消除的污染積累區域。

分層土壤在垂直方向上的差異較為明顯，不同土層在質地、孔隙結構和有機質含量方面存在較大不同。表層耕作土中有機質和黏土礦物含量較高，具有較強的陽離子交換能力，因此對部分重金屬有較強的吸附作用。中層和下層土壤多由粉砂或砂質材料組成，其滲透性較好，孔隙之間連通性較強，容易形成重金屬快速向下移動的通道。由於不同土層之間孔隙率變化較大，同時水流條件發生改變，在層間介面處常出現局部積水或指狀水流現象，這些條件為重金屬在不同土層之間的擴散提供了動力和通道^[2]。

目前關於分層土壤中重金屬遷移的研究，多集中在重金屬含量變化和隨水流移動的過程，對土層之間擴散過程的研究仍然不足。現有研究中常將土壤視為結構均一的介質，這種處理方式難以反映真實土壤剖面中層次差異與降雨作用共同影響下的遷移過程。在水流條件、化

學環境以及吸附和釋放過程等多種因素共同作用下，重金屬在土壤剖面中的分佈和變化具有明顯的時間和空間差異，這也增加了污染預測和治理工作的難度。

1 分層土壤結構與水動力過程

分層土壤的形成與母質類型、地形條件、氣候變化以及人類活動的長期作用有關，在垂直方向上表現出顆粒組成、孔隙結構和水流特性的明顯差異。表層土壤多為耕作土或受人為影響的土壤，質地以壤土為主，孔隙較多，有機質和黏粒含量較高，因此具有較強的保水能力和陽離子交換能力。中層土壤通常為粉砂或黏土層，結構較為緊密，孔隙較小且分佈不均，水分通過能力較弱。下層多為砂質母質或風化岩層，孔隙之間連通性較差，水分滲透速度明顯降低。這種垂直方向上的差異，決定了土壤對降雨入滲和污染物移動的基本反應方式。

在降雨過程中，表層土壤會迅速吸收水分，並形成較高的含水區，水分在毛細作用下逐漸向下移動。當降雨強度和持續時間超過土壤的入滲能力時，部分水分會沿著較大的孔隙、生物通道或土壤裂隙快速下滲，繞過中層滯水區域，在短時間內形成較大的水流量。這一過程會縮短污染物在土壤中的停留時間，減弱表層土壤的吸附作用，從而增加重金屬向下移動的可能性。

隨著入滲過程繼續，中層土壤較低的滲透能力會使水分在其上方暫時積聚，形成局部滯留區域。在該區域內，溶質濃度逐漸升高，使重金屬更容易與土壤顆粒發生吸附反應，一部分污染物由溶解狀態轉變為吸附狀態。水分滯留時間越長，這種吸附過程越明顯。降雨結束後，殘留水分會繼續緩慢下滲，在剖面中形成相對穩定的含水區，並為後續由濃度差引起的擴散過程提供持續推動力。

在非飽和條件下，不同土層之間孔隙率和滲透能力的差異容易引發指狀水流現象，局部形成狹窄的水流通路，使污染物繞過土壤基質區，快速進入更深層土壤。這類優先通道在強降雨或多次幹濕變化過程中更加明顯，會加重污染物分佈的不均勻性，並增加其在深層土壤

中的累積深度，是重金屬層間擴散的重要水流條件。

土壤中的水分變化還受到降雨和乾旱交替過程的影響。乾燥階段會增強土壤吸水能力，提高初始入滲速度，而濕潤階段則會加強溶質的釋放和重新分佈，推動污染物在不同形態之間發生轉變，並繼續向下移動^[3]。在長期反復作用下，這種交替過程會產生明顯的累積效應。

2 降雨驅動下重金屬遷移機理

降雨不僅為土壤不斷補充水分，還推動了重金屬在土壤剖面中的分階段移動過程。它的影響不僅體現在降雨發生時的短時間水流運輸上，還體現在降雨前後對孔隙水流狀態和化學條件的改變，從而長期影響重金屬的存在形式和在土體中的分佈情況。

在降雨開始階段，表層土壤會迅速吸收水分，並逐漸接近飽和狀態。隨著孔隙水壓力上升，溶解態和絡合態重金屬會隨水分沿垂直方向發生對流移動。降雨強度直接影響水分下滲速度和水流大小。強降雨會形成較大的水流通量，使污染物在短時間內穿過表層的吸附區域，並進入中層或更深的土壤。較大的孔隙和裂隙通道容易形成優先水流，這種水流使土壤中的水分分佈更加不均，也明顯加快了重金屬的移動速度，使一部分污染物繞過滯留區域，在較短時間內直接進入深層土壤。

隨著雨水繼續向下滲透，中層土壤由於滲透能力較弱，水分容易在其上方積聚，使污染物在局部含水區域內不斷集中，濃度逐漸升高。在這一階段，控制遷移的主要過程不再是快速對流，而是吸附、解吸以及重新分配過程。重金屬會在土壤顆粒表面不斷發生吸附與釋放，形成變化中的平衡狀態^[3]。土壤的酸鹼程度、陽離子交換能力以及有機質含量等條件，會明顯影響吸附速度和吸附數量。一部分溶解態污染物被土壤固相吸附後，其可移動比例會降低；而在反復幹濕變化或離子交換作用下，已經吸附的重金屬又可能重新釋放，恢復繼續移動的能力。

在降雨結束之後，水分向下移動的速度明顯減慢。此時，表層與中層之間由於污染物濃

度不同，會形成清晰的濃度差。隨著水流作用減弱，分子擴散逐漸成為主要的遷移方式。重金屬會從濃度較高的區域緩慢向濃度較低的區域擴散。雖然這一過程速度較慢，但持續時間較長，是導致重金屬在深層土壤中長期累積的重要原因。濃度差引起的擴散過程通常可以用菲克第一定律進行描述：

$$J = -D \frac{\partial C}{\partial z}$$

式中， J 為擴散通量， D 為有效擴散係數， $\partial C/\partial z$ 為濃度梯度。 D 值受土壤孔隙結構、含水率與溫度等因素影響，層間孔隙率突變造成擴散不連續，影響通量連續性。

除了上述主導機制，重金屬在土壤中的遷移形態也在降雨過程中發生變化。水溶態和弱絡合態污染物最易遷移，隨對流與擴散進入深層^[4]；而顆粒結合態和沉澱態在一定條件下可發生解吸或溶解，成為潛在的再釋放源。水溶態和弱絡合態污染物最易遷移，隨對流與擴散進入深層；而顆粒結合態和沉澱態在 pH 下降和鹽分稀釋等條件下可發生解吸或溶解，成為潛在的再釋放源^[5]。

在這一遷移過程中，非飽和區中的指狀水流現象尤為重要。在土壤內部，由於不同位置的水流能力存在差異，降雨入滲時往往不會形成均勻下滲，而是在剖面中出現不規則的指狀水流通道。這些水流通道阻力較小，為污染物提供了更容易通過的路徑。重金屬可以沿著這些通道繞過部分吸附能力較強的土壤區域，直接進入下部土層，從而明顯加快其向下移動的速度，並增加在深層中的累積程度。指狀流並不是穩定存在的，其形成和變化與土壤結構條件、降雨強弱以及降雨前土壤的含水狀態密切相關。在降雨強度較大或土壤經歷多次幹濕變化時，這類水流通道更容易出現，其對重金屬遷移路徑和遷移深度的影響也更加明顯。因此，在建立相關模型時，有必要對指狀流過程進行明確描述，以反映其在污染物遷移中的實際作用^[6]。

3 層間擴散的理論機制模型

層間擴散是分層土壤中重金屬移動的重要

方式。在不同土層之間，由於重金屬濃度存在差異，加上孔隙結構不一致以及水流條件不同，溶解態重金屬會在這些因素共同作用下，持續跨越土層介面向下或向上移動。這一過程通常發生在水流速度較低或降雨結束之後，但其作用時間較長，對重金屬在剖面中的分佈具有持續影響。為了描述這種隨水流和擴散共同發生的遷移過程，研究中常採用對流—擴散方程，即 Advection-Dispersion Equation (ADE)。在實際分析中，多使用其一維形式來刻畫重金屬沿垂直方向的遷移行為，其基本表達形式為：

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial z^2} - v \frac{\partial C}{\partial z} - \rho_b \frac{\partial S}{\partial t}$$

其中：

- C 為孔隙水中重金屬濃度 (mg/L)
- D 為總彌散係數，包含機械彌散項 αv 與分子擴散係數 D_m ，即 $D = D_m + \alpha v$
- v 為垂向平均對流速度 (cm/s)
- ρ_b 為土壤體積密度 (g/cm³)
- S 為單位土壤品質吸附態重金屬濃度 (mg/g)
- z 為深度 (cm)
- t 為時間 (s)

ADE 模型認為，溶質在土壤中的移動過程符合 Fick 定律，並假定水分在土體中以連續、均勻的方式流動。這種模型通常適用於水流條件較為穩定、孔隙結構差異不大、且未出現明顯優先水流的理想情況。然而，在垂直方向差異明顯的分層土壤中，這些前提條件往往難以滿足。由於不同土層在結構和水流條件上存在明顯變化，ADE 模型在描述實際遷移路徑和污染物空間分佈時，容易產生一定偏差。

彌散係數 D 的確定需結合實測土壤孔隙率與流速條件，通過穿透曲線反演或室內土柱試驗獲得。彌散長度 α 通常取決於土壤質地與結構變化程度，一般砂壤土中取值為 1 - 10 cm，而黏性土壤則更低。分子擴散係數 D_m 與重金屬種類和孔隙水溫度相關，通常為 $10^{-5} \sim 10^{-6} \text{cm}^2/\text{s}$ 。

吸附—解吸過程對遷移能力具有強調控作用，常採用 Langmuir 或 Freundlich 等溫吸附模

型建模。Langmuir 模型適用於單層飽和吸附，表達為：

$$S = \frac{K_L C}{1 + a_L C}$$

其中 K_L 為最大吸附容量 (L/g)， a_L 為平衡常數。Freundlich 模型適用於非理想表面吸附，運算式為：

$$S = K_F C^n$$

其中 K_F 、 n 為經驗常數，適合描述非線性吸附行為。二者的選擇應基於等溫吸附實驗結果。

ADE 模型在預測濃度場、評估累積深度方面具備一定準確性，但在實際剖面中，降雨引發的指狀流、裂隙通道、突變孔隙率等會造成流線偏轉與局部富集，模型低估污染物垂向遷移範圍。

為克服 ADE 在非均質條件下的不足，可引入雙孔隙模型 (dual-porosity model) 或雙域模型 (dual-permeability model)，考慮快慢流區之間的溶質交換。雙域模型的形式如下：

$$\frac{\partial C_f}{\partial t} = D_f \frac{\partial^2 C_f}{\partial z^2} - v_f \frac{\partial C_f}{\partial z} - \rho_b \frac{\partial S_f}{\partial t} - \lambda (C_f - C_s)$$

$$\frac{\partial C_s}{\partial t} = -\lambda (C_s - C_f)$$

其中 C_f 、 C_s 分別為快流域（如指狀流）與慢流域（基質區）中的濃度， λ 為交換係數，反映兩區域間濃度交換速率。該模型通過設定不同的流速和擴散參數，更貼近自然條件下的擴散行為。

指狀流在剖面中分佈隨機，模型中可通過設定高滲透性隨機場或嵌套網格方法加以模擬。數值類比結果顯示，指狀流對總遷移通量貢獻可達 30% 以上，是 ADE 無法捕捉的重要非均質結構效應。

模型參數的設定可基於實地土壤剖面的分層取樣結果，並結合溶質穿透試驗獲取基礎資料。在此基礎上，可通過反演和參數調整方法，對實驗資料進行擬合，從而得到更符合實際情況的遷移參數組合。在不同降雨強度和持續時間條件下開展模型計算，有助於觀察土層介面

區域對遷移過程的回應變化，並判斷哪些區域對水分和污染物移動更加敏感。這些結果可以為污染阻隔位置的選擇以及採樣方案的調整提供直接參考^[7]。

4 環境風險與防控啟示

在分層土壤中，降雨推動的重金屬層間擴散不僅會加快污染物向下移動，還會明顯增加深層土壤和地下水受到污染的風險。在長期反復的降雨和乾旱過程中，表層土壤中積累的污染物會在多次水流作用和擴散作用下，逐步穿過中層阻滯區域，進入下部土壤，部分情況下甚至會到達潛水面。由於深層土壤通常孔隙較少、取樣和處理較為困難，一旦形成穩定的污染累積區，後期治理難度和成本都會明顯提高，所需治理時間也往往遠長於表層污染。

深層重金屬污染在土壤剖面中具有較強的隱蔽性，常規的地表取樣方式很難及時發現其不斷增加的趨勢^[8]。當重金屬在深層含水區內發生重新分佈並不斷聚集時，可能形成持續釋放的污染源，對淺層地下水水質造成長期影響。在向下移動過程中，高濃度的孔隙水污染物往往會沿著裂隙或指狀水流通道進一步加快擴散速度，使污染在空間上的分佈更加不均，這也增加了監測和預警工作的難度。

在農業土壤環境中，降雨引起的層間擴散過程可能導致耕作層以下、作物根系分佈區域內的重金屬含量不斷升高，從而提高農作物吸收污染物以及通過食物鏈積累的風險^[9]。水溶態或弱絡合態的重金屬可以隨著根區水分被植物吸收，並進入地上部分組織，增加農產品超標的潛在隱患。隨著時間推移，這種由降雨引起的擴散和累積過程會逐漸改變土壤原有的生態功能，對微生物數量、群落結構以及土壤酶活性產生影響，從而降低整體土壤品質和健康水準。

在應對這些風險時，防控措施需要從多個層面同時展開。污染源控制是首要環節，應加強對含重金屬廢水、廢渣以及農業投入品使用的管理，儘量減少新的污染物進入土壤^[10]。對於已經受到污染的區域，應結合土壤分層結構開展更為細緻的調查，重點關注中層水分滯留區

域和深層可能出現的污染積累區。通過多深度剖面取樣和孔隙水監測，可以更清楚地識別污染物的遷移路線以及指狀水流通道的位置。

在防止污染繼續擴散的技術措施上，改善表層土壤的理化條件、提高其吸附和固定重金屬的能力，是降低向下遷移風險的重要方式。施加礦物類固化材料或有機改良材料，可以增強重金屬在表層土壤中的穩定性，減少其以溶解狀態存在的比例。在中層或不同土層交界處設置阻隔層或高吸附層，如黏土墊層、膨潤土墊層或生物炭層，有助於攔截向下移動的溶解態污染物。同時，通過調整農田排水方式，分散雨水集中流入，降低短時間內的人滲強度，也可以在一定程度上減弱優先水流的形成。

在風險評估和監測工作中，需要特別關注時間變化和空間差異。針對降雨事件後可能出現的短時間高強度遷移，應在降雨前後適當提高監測頻率，以捕捉污染物快速增加的過程。在長期幹濕反復變化的條件下，還應持續跟蹤土壤中濃度差的變化情況以及孔隙水中溶質的累積過程，為後續修復措施的調整提供可靠的資料支援。

通過多維度防控措施的聯合應用，分層土壤中重金屬層間擴散風險可以在源頭削弱和路徑阻斷兩個層面得到有效管控，降低深層累積和地下水污染概率，為農業生態安全和土壤環境品質的長期保障提供技術保障。

5 結論

在分層土壤中，降雨推動的重金屬遷移過程表現出多種作用同時發生，並且在不同階段具有明顯變化。在降雨入滲過程中，不同土層的孔隙結構、水分變化情況以及化學條件共同影響污染物在垂直剖面中的移動方式和分佈位置。在降雨開始階段，對流作用占主導地位，水分迅速向下流動，推動重金屬快速穿過表層污染區域。隨著水分在中層土壤或土層交界處停留，污染物逐漸集中，形成局部濃度較高的區域，這些區域隨後成為擴散過程的重要來源。在降雨間歇期以及反復的幹濕變化過程中，由濃度差引起的緩慢擴散持續發生，並在較長時間內主導遷移過程，使重金屬逐步在深層土壤

中形成潛在的累積區域。指狀流等非均質水動力路徑在不飽和狀態下增強遷移不確定性，是污染深層突破的重要通道。

理論建模採用對流—彌散—吸附方程（ADE）為基礎框架，融合 Langmuir 與 Freundlich 等溫吸附模型，構建了適應層間結構特徵的遷移類比體系^[11]。模型參數如彌散係數、平均流速、土壤體積密度與最大吸附容量係數需根據實驗測定或田間監測資料進行分層賦值，確保對非均質性的回應能力。類比結果顯示，污染物在中層土壤及不同土層交界位置常出現移動變慢並停留的情況。參數敏感性分析進一步表明，吸附過程的變化速度以及孔隙結構在介面處的突然變化，是影響污染物向下移動深度和累積程度的主要因素。傳統的 ADE 模型在一定程度上可以解釋污染物的整體遷移趨勢，但在存在指狀水流、裂隙滲流以及快慢水流區域相互交換等情況時，其適用範圍明顯受限。在這些

非線性遷移條件下，模型往往難以準確反映真實的遷移路徑和分佈特徵，因此需要通過引入雙重孔隙模型或採用分區參數設定的方法，對模型進行擴展和調整，以提高模擬結果的可靠性和精度。

環境風險評估結果表明，降雨推動的層間擴散過程會加快污染物向深層土壤和潛水層的移動，使污染更加隱蔽，同時顯著提高後續治理所需的時間和成本^[12]。一旦深層污染累積形成，常規修復措施往往難以有效到達污染區域，其釋放的污染物還可能持續進入地下水系統，造成長期影響。在農業生態環境中，重金屬的垂向移動還可能穿過作物根系分佈區域，增加植物吸收污染物的可能性，並提高食物鏈暴露風險。在應對措施上，應加強污染源頭的控制，並結合表層固定技術，同時設置具有較強攔截能力的介面層和合理的排水調控結構，以減少降雨條件下形成的高強度遷移通道。

參考文獻：

- [1] 周國華. 土壤重金屬生物有效性研究進展[J]. 物探與化探, 2014, 38(6): 1097-1106.
- [2] 郭超. 雨水花園集中入滲對土壤和地下水影響的試驗研究[D]. 陝西:西安理工大學, 2019.
- [3] 陳其鮮, 楊琴, 任書生. 拔節期水分虧缺對沙地農田玉米光合特性及物質積累的影響[J]. 水土保持通報, 2015, 35(04): 91-96.
- [4] 陳蘇, 孫麗娜, 晁雷, 等. 不同濃度組合的鎘、鉛在不同污染負荷土壤中的吸附-解吸動力學行為[J]. 應用基礎與工程科學學報, 2007, 15(1): 32-44.
- [5] SEMPLE K T, DOICK K J, WICK L Y, HARMS H. Microbial interactions with organic contaminants in soil: Definitions, processes and measurement[J]. Environmental Pollution, 2007, 153(1): 119-126.
- [6] ALLOWAY B J. Heavy metals in soils: Trace metals and metalloids in soils and their bioavailability[M]. 3rd ed. Dordrecht: Springer, 2013.
- [7] 鄭順安. 我國典型農田土壤中重金屬的轉化與遷移特徵研究[D]. 浙江: 浙江大學, 2010.
- [8] 陳璐, 文方, 程豔, 等. 鉛鋅尾礦庫周邊土壤重金屬污染特徵及環境風險[J]. 中國環境監測, 2017, 33(1): 82-87.
- [9] 蘇耀明, 陳志良, 雷國建, 等. 多金屬礦區土壤重金屬垂向污染特徵及風險評估[J]. 生態環境學報, 2016, 25(1): 130-134.
- [10] 高永華, 王金, 趙莉, 等. 汙灌區土壤-植物系統中重金屬分佈與遷移轉化特徵研究[J]. 河北農業大學學報, 2006, 29(5): 52-56.
- [11] 李秋華, 夏品華, 吳紅, 等. 用物理-生態集成技術局部控制富營養化[J]. 環境科學, 2012, 33(07): 2303-2308.

[12] 陳棟為, 陳曉宏, 孔蘭. 基於生態足跡法的區域水資源生態承载力計算與評價——以珠海市為例[J]. 生態環境學報, 2009, 18(6): 2224-2229.

版權聲明

© 2025 作者版權所有。本文依據“知識共用署名 4.0 國際授權合約”(CC BY 4.0)以開放獲取方式發佈。該許可允許使用者在任何媒介中自由使用、複製、傳播與改編文章(含商業用途),惟須明確署名原作者及出處,並注明所作修改(如有)。完整協議詳見:<https://creativecommons.org/licenses/by/4.0/deed.zh-hans>

出版聲明

所有出版物中的陳述、觀點及資料僅代表作者及供稿者個人立場,與 Brilliance Publishing Limited 及/或編輯人員無關。Brilliance Publishing Limited 及/或編輯人員對因內容所提及的任何理念、方法、說明或產品所導致的人身或財產損害概不負責。

DOI: 10.53104/xdkxsts.2025.01.02.004

推薦系統中演算法強化與使用者偏好共演化機制研究

周波¹

1. 渤海大學，遼寧 錦州，121013

摘要：推薦系統是智慧資訊分發中的關鍵技術，已經成為人們獲取資訊方式中的重要部分。在實際運行中，系統會根據使用者的點擊、流覽和停留等行為反復調整模型參數，這種調整提高了個性化程度和推薦準確性，但同時也在慢慢改變使用者的興趣結構和偏好取向。本文從運行機制出發，討論演算法強化過程與使用者偏好變化之間相互影響、相互塑造的關係，研究首先介紹了推薦系統的基本技術結構和演算法強化的工作方式，指出使用者行為在系統中具有雙重意義，用戶行為直接表達個人興趣和選擇傾向和這些行為會被系統記錄並作為回饋資料重新輸入模型，用來指導下一輪推薦，兩種作用在系統中同時存在，使演算法學習和使用者選擇形成一個持續運轉的反饋回路。文章從機制層面分析了使用者偏好在推薦系統中的變化過程，說明在長期重複的回饋作用下，用戶更容易反復接觸相似內容，興趣逐漸集中在有限的內容類別型上，而其他類型被看到的機會不斷減少，這種偏好收縮和多樣性下降，並不是由個體主觀意願單獨決定的，而是由系統回饋結構本身推動形成的結果。本文提出一個演算法與使用者共同變化的系統模型，將推薦系統理解為一個由演算法學習和人類行為共同驅動的整體系統，這一系統在持續回饋過程中不斷調整自身狀態，表現出明顯的適應性、不均勻變化以及對早期選擇高度依賴的特徵。基於這一認識，文章從工程設計角度討論了強化目標設定、系統穩定性以及內容多樣性等問題，強調在系統設計階段應正視演算法強化對系統長期走向的影響。本文從系統運行和機制結構的角度解釋推薦系統在長期使用中的行為邏輯，為理解演算法強化條件下的人與演算法互動關係提供一種清晰、直接的分析思路。

關鍵字：推薦系統；演算法強化；用戶偏好；反饋回路；共演化機制；系統演化

Research on the Co-Evolution Mechanism of Algorithmic Reinforcement and User Preferences in Recommender Systems

ZHOU Bo¹

1. Bohai University, Jinzhou 121013, P.R.China

Correspondence to: ZHOU Bo; Email: bo.1996.85@outlook.com

Abstract: The recommendation system, as a core technology for intelligent information distribution, has become an important component of the human information interaction structure. With the continuous development of the algorithm reinforcement mechanism, the system continuously optimizes the model performance through user behavior feedback, achieving personalized and high-precision recommendations while gradually influencing and reshaping users' interests and preferences. This paper, from a mechanism perspective, explores the co-evolution

收稿日期：2025-12-10 返修日期：2025-12-30 錄用日期：2026-01-09 出版日期：2026-01-13

通信作者：bo.1996.85@outlook.com

引用格式：周波. 推薦系統中演算法強化與使用者偏好共演化機制研究[J]. 現代科學探索, 2025, 1(2): 37-48.

relationship between algorithm reinforcement and user preferences. The study first analyzes the technical structure of the recommendation system and the logic of algorithm reinforcement, pointing out that user behavior in the system plays a dual role of expressing preferences and generating feedback, thereby forming a continuous reinforcement loop. Subsequently, through mechanism analysis, it reveals the evolution path of user preferences in the recommendation system and demonstrates the internal logic of preference convergence and diversity decline caused by the long-term feedback cycle. Further, this paper proposes a system model of algorithm and user co-evolution, arguing that the recommendation system can be regarded as a self-organizing system driven by both algorithm learning and human behavior. This system exhibits complex characteristics such as adaptability, non-linearity, and path dependence in dynamic feedback. Based on this, this paper conducts mechanism-based reflections on the goal design of reinforcement, system stability, and diversity issues from the engineering perspective, emphasizing that attention should be paid to the shaping effect of algorithm reinforcement on the long-term evolution of the system during the design stage of the recommendation system. The research of this paper does not involve algorithm effect evaluation or empirical verification, but is dedicated to explaining the long-term behavioral logic of the recommendation system from the perspective of system evolution and mechanism, providing theoretical frameworks and engineering inspirations for understanding the human-algorithm interaction relationship under algorithm reinforcement.

Key words: recommender system; algorithmic reinforcement; user preferences; feedback loop; co-evolution mechanism; system evolution

引言

推薦系統已經滲透到人類日常資訊消費的幾乎所有場景。它們在新聞分發、短視頻推薦、社交媒體資訊流、線上教育以及電子商務中承擔著內容篩選與個性化呈現的關鍵角色。面對爆炸式增長的信息量，推薦系統通過演算法自動化決策過程，替代了用戶的自主篩選行為，從而成為數位環境中人機交互的核心機制^[1]。人們在流覽、點擊、評論、分享的過程中不斷生成行為資料，而這些資料又被系統捕獲、分析並用於演算法的持續優化。推薦系統因此不僅僅是資訊分發的工具，也成為塑造資訊接觸結構的重要力量。

在技術邏輯層面，當前主流的推薦演算法大多依賴使用者的歷史行為進行建模。系統通過收集點擊、流覽、停留時間、購買記錄等回饋信號，構建個體化的興趣畫像，並在此基礎上優化模型參數，以實現更高的預測精度。這一機制具有自學習特徵，模型能夠隨著使用者行為的累積而動態更新，從而在不斷的反覆運算中提升匹配度和用戶滿意度。工程實踐中，這種基於回饋的演算法優化被視為智慧化推薦的

核心優勢。

這一過程並非單向的識別與回應關係。演算法在根據使用者行為強化自身性能的同時，也在持續塑造用戶的注意力與偏好。用戶所表現出的興趣往往並非完全獨立的主觀選擇，而是在演算法生成的內容環境中被引導與限定的結果。使用者行為成為演算法學習的輸入，而演算法輸出又反過來影響使用者的行為取向。演算法優化與使用者回應之間形成了一個不斷迴圈的回饋系統，使得個體的興趣軌跡與演算法的演化路徑交織在一起^[2]。

這一迴圈帶來了一個值得深入思考的問題：在演算法的持續強化過程中，用戶的偏好是否仍然可以被視為自主形成的心理結構？或者說，使用者的興趣是否正被演算法所生成的推薦邏輯系統性地重塑？推薦系統的優化目標通常聚焦於提高交互頻率與平臺留存率，用戶偏好的生成則被隱含地嵌入到這一優化框架中。當演算法強化與使用者行為相互作用時，原本應作為系統輸入的使用者偏好，逐漸轉化為演算法輸出的一部分。演算法不再只是識別興趣，而是在重複的回饋迴圈中生產興趣、放大興趣、固化興趣。

在這一意義上，推薦系統已經超越了單純的技術工具屬性，成為一個具有自組織特徵的社會—技術系統。它在運行中不斷學習、調整、適應，而用戶的認知、情感和行為也在系統的動態輸出下逐步演化。演算法強化的技術邏輯與人類心理機制交織，使得個性化推薦演化為一種共生的結構性過程。系統不再是中立的媒介，而是具有方向性的行為塑造力量。

本文正是基於這一現象展開機制層面的探討。研究不討論推薦效果的優劣，也不進行實驗驗證，而是聚焦於演算法強化與使用者偏好的互動關係。通過分析推薦系統的結構邏輯與強化機制，揭示演算法在優化過程中如何影響使用者偏好的形成路徑，進而探討人—演算法系統在長期交互中呈現出的共演化特徵。本文試圖從系統演化的視角理解這一機制，為未來關於推薦系統長期行為的研究提供理論基礎。

2 推薦系統與演算法強化的技術基礎

2.1 推薦系統的基本結構

推薦系統是一種通過演算法模型對使用者與內容進行匹配的智慧資訊過濾機制，其目標是在龐大的資訊集合中為特定使用者提供最可能感興趣的項目。在資訊超載的數位環境中，推薦系統通過分析歷史行為資料構建使用者偏好模型，以減少使用者的搜索成本並提升個性化服務品質^[3]。推薦系統的運行依賴於四個基本組成部分：使用者、內容或物品、行為資料與推薦模型。這四個部分構成了一個動態交互的閉環，使系統能夠在不斷地資料流程動中實現自學習與性能優化。

使用者是推薦系統的起點與核心。系統需要在大量使用者群體中識別個體特徵，從而生成差異化的推薦結果。使用者在系統中的存在並不僅僅是被動的接收者，他們通過點擊、流覽、收藏、評論、購買等行為不斷提供資訊信號，這些信號成為演算法優化的直接依據^[4]。使用者的行為構成了推薦系統的感知層，決定了系統能夠捕捉到怎樣的興趣輪廓與偏好趨勢。

內容或物品是推薦系統的另一端。系統需

要對內容資源進行結構化處理與特徵表示，使其可以被演算法識別與計算。在實際系統中，無論是視頻、文章還是商品，系統通常會通過特徵提取技術將其轉化為向量化的資料結構，以便與使用者興趣空間進行匹配。推薦系統不僅處理內容的顯性資訊，還通過隱語義建模方式捕捉潛在屬性，從而發現抽象層面的偏好模式^[5]。

行為資料在系統中承擔橋樑作用。它連接使用者與內容，是系統學習與回饋的基礎。資料的收集通常通過埋點、日誌記錄和使用者交互介面完成。點擊、停留時間、跳出率、點贊次數等指標構成了系統的輸入信號。資料經過清洗、預處理與建模後進入推薦模型的訓練階段，成為演算法優化的核心資源。資料品質直接影響模型性能，而模型對資料的依賴也導致推薦系統具有明顯的歷史依賴特徵^[4]。

推薦模型是整個系統的智慧中樞，其核心任務是從資料中學習模式與規律，形成可泛化的預測能力。早期推薦系統主要基於協同過濾思想，即通過比較用戶或物品之間的相似性實現推薦。隨著機器學習技術的發展，基於矩陣分解、深度神經網路等方法逐漸成為主流^[3]。這些模型能夠從大規模行為資料中提取高維特徵，並在非線性空間中進行匹配預測，從而提高推薦效果。

推薦系統的運行邏輯可以概括為一個動態的閉環。使用者的行為首先被採集並轉化為結構化資料，系統利用這些資料訓練模型，模型生成新的推薦結果，用戶在接收推薦後再次產生新的行為，而這些新的行為又被系統採集進入下一輪學習。這種迴圈形成了自我調整的資料流程動機制，系統性能的提升依賴於回饋品質，而回饋品質又受到系統輸出的影響^{[3][5]}。用戶的每一次點擊都可能在無形中影響模型的未來更新方向，使演算法在演化過程中不斷重構用戶興趣空間。

在實際應用中，推薦系統常常面臨多目標優化問題。平臺可能同時關注點擊率、轉化率、停留時間與用戶滿意度等指標。不同目標之間存在潛在的張力，系統需要在短期收益與長期用戶體驗之間取得平衡。演算法在此過程中會形成複雜的優化路徑，這種多維度的學習與調

節進一步加深了推薦系統的動態複雜性^[6]。

2.2 演算法強化的基本邏輯

演算法強化指系統在運行過程中通過回饋信號實現性能的自我改進。這一概念源自強化學習的思想，即智慧體在環境中通過行動與回饋不斷學習最優策略。在推薦系統中，演算法強化表現為模型利用使用者的行為回饋來調整推薦策略，使得系統在連續交互中逐漸趨向於最能引發積極反應的推薦模式^[7]。

推薦系統的回饋信號通常以點擊、停留時長、轉化行為或互動頻率等形式存在。這些回饋被視為系統對某一推薦結果的回報評估。演算法的優化目標通常是最大化某種回報函數，代表系統希望達到的行為效果。例如在短視頻平臺中，演算法可能以使用者的觀看時長作為主要回報信號，在電商平臺中則可能以購買轉化率為主要優化指標。回報函數的設定直接決定了演算法強化的方向，也隱含地規定了系統的演化軌跡^[7]。

演算法強化過程可以理解為一個持續的決策鏈條。每一次推薦都是演算法在當前狀態下對未來行為的一種預測與嘗試。系統根據使用者的即時回饋調整模型參數，使後續推薦更符合預期。強化學習在此過程中提供了一種動態平衡機制，使系統能夠在探索新內容與利用已有規律之間尋找最佳策略^[8]。

強化過程具有顯著的路徑依賴性。推薦系統初期的行為資料往往在演算法中佔據不成比例的權重，這些早期資料構成了系統的“經驗基礎”。模型在訓練中會逐漸偏向於這些早期形成的偏好模式，導致演算法在後續階段更傾向於重複強化既有的興趣結構。這種依賴性使系統在長期運行中可能出現興趣收斂與推薦趨同的現象^{[5][7]}。

演算法強化不僅是一種技術過程，也是一種認知生成過程。使用者行為既是演算法學習的依據，也是被演算法塑造的產物。使用者在演算法推薦的內容中進行選擇，這些選擇又反過來驗證演算法的判斷。演算法因此獲得“正回饋”，進一步強化相似的內容分發模式。推薦系統逐步形成自迴圈的結構，使演算法強化與使用者回應之間呈現相互放大的趨勢^{[2][5][8]}。

在工程實踐中，演算法強化通常通過線上學習機制實現。系統會在每一次使用者交互後即時更新部分模型參數，使得推薦策略能夠適應環境變化。這種更新方式使系統具備高度靈活性，但也帶來動態不穩定的風險。當系統對短期回饋過度敏感時，演算法可能出現頻繁的策略波動，導致推薦品質下降或用戶體驗斷裂。為緩解這種問題，部分系統會引入延遲回饋建模機制，通過時間視窗平滑參數更新，降低短期雜訊的影響^{[6][7]}。

從系統演化的角度看，演算法強化構成了推薦系統的內在驅動力。系統通過強化機制不斷優化自身行為模式，實現自組織與適應性演化。每一次回饋更新都意味著系統的內部狀態發生微小調整，積累到宏觀層面便可能引發系統結構的整體偏移。演算法因此不再是被動響應機制，而是具有主動演化特徵的適應性實體^[8]。

3 演算法強化驅動的使用者偏好演化機制

推薦系統的運行不僅是演算法對資料的計算過程，也是一種社會—技術層面的交互過程。演算法強化使得推薦系統在技術上實現自我優化的同時，在內容分發過程中參與了用戶注意力與偏好的塑造。這種交互表明用戶興趣、認知與選擇習慣並非孤立存在，而是在演算法模型與回饋迴圈中持續演化，這與“使用者與演算法共演化”的結構性機制一致^{[9][10]}。

3.1 使用者行為作為強化信號的機制分析

在推薦系統中，使用者行為資料被視為演算法優化的核心資源。點擊、收藏、停留時長、點贊、評論等行為都被設計為回饋信號，直接影響模型的更新方向。系統在獲取這些信號時往往假定它們真實反映了用戶偏好。然而，從系統機制角度看，使用者行為不是完全獨立的心理意圖表達，而是在演算法推薦環境中產生的反應結果。推薦系統通過反覆展示演算法認為高價值的內容，使使用者在有限的內容集合中做出選擇，這一過程引發了回饋信號的“約束性表達”^{[9][10]}。

這種限定性表達構成了使用者偏好的半結構化資料空間。演算法在這一空間中學習興趣分佈，但這種興趣空間本身已受到既有推薦策略的影響。當使用者行為被用作訓練信號時，系統在優化模型的同時也在強化自身對這一行為空間的依賴，從而使行為資料既是輸入又是約束機制^{[10][11]}。這種編碼與過濾作用在機器學習中常見於對輸入特徵的精煉，在推薦系統中體現為對“強信號”和“弱信號”的不同解讀與放大。

推薦系統的回饋邏輯改變了傳統資訊選擇的路徑，使“被動暴露”逐漸替代了“主動搜索”。推薦演算法根據使用者與候選內容的歷史行為分配注意力，而使用者在演算法分配的注意力空間中做出的每一次點擊，都被量化為回饋信號^{[10][12]}。隨著時間累積，這種迴圈不斷削弱了使用者在資訊選擇中的自主性，使得使用者偏好受到演算法結構的強約束。

從心理機制層面分析，用戶行為本身帶有情緒與動機成分。演算法推薦常常通過介面設計、排序策略等方式放大內容的情緒吸引力，使用戶在愉悅或好奇等心理狀態下產生更多高回饋行為。系統捕捉到這些信號後進一步放大同類內容，使使用者形成習慣性偏好^{[9][10]}。這種情感驅動的行為強化與演算法參數更新同步進行，構成了技術層面與心理層面的共振。

行為信號的可計算性使演算法能夠對使用者偏好進行有效估計，但這種簡化也伴隨資訊損失。複雜的心理興趣被折疊成可統計指標，在高維空間中被模型平滑為穩定趨勢^{[10][11]}。這種模型追求穩定性的策略往往優先強化重複性行為，而忽視偏好本身的流動特性，在長期演化中可能造成興趣結構的漸進固化。

3.2 強化反饋回路的形成過程

推薦系統的回饋結構可抽象為一個閉環過程：初始偏好 → 推薦內容 → 使用者回應 → 模型強化 → 下一輪推薦。這一迴圈不僅是技術流程，也是動態認知生成機制。系統在每一輪迴圈中吸收使用者回饋信號，調整內部權重，生成更新的推薦內容；使用者在新環境中反應再次形成資料登錄^{[10][11]}。

初始偏好通常來自用戶顯性輸入或冷開機

特徵，如註冊興趣標籤或早期行為歷史。基於此資訊，系統生成第一輪推薦。使用者對推薦內容的回應成為演算法學習的第一批回饋資料，而這些資料在初期階段對模型方向具有顯著影響，因為初期資料量少但權重往往較大^{[9][10]}。這種早期效應常常導致參數空間快速收斂到特定模式。

隨著系統運行，推薦內容逐步趨向高回饋區域。演算法通過點擊率等指標判定哪些內容能獲得更高使用者回應，並提高其曝光優先順序^[14]。這種機制在理論上能夠快速提升短期回饋指標，但長期來看卻可能壓縮用戶的探索空間，使推薦策略陷入“高回饋區域迴圈”。

用戶回應是閉環中的關鍵節點。演算法將使用者行為視為環境回饋，而模型通過梯度調整等機制更新參數。這種基於統計規律的強化方式無法區分使用者的真實興趣和環境誘導行為^{[10][11]}。由於演算法並不理解使用者內在心理意圖，其“學習”實際上是對行為模式的概率化估計，這種估計在長期迴圈中可能產生系統性偏移。

在進入下一輪推薦後，新策略往往更偏向歷史高回饋區域，使得使用者接收到的內容結構更加集中，從而引發行為模式趨同^{[10][12]}。短期內這似乎提高了預測準確度，長期則會使系統多樣性下降、興趣空間壓縮。這種結構化反饋回路與認知層面的習慣形成高度耦合，使推薦系統具有自組織特性。

介面設計與交互機制在強化回路中起到輔助作用。推薦系統通過視覺佈局、推薦頻率等方式控制使用者注意力路徑，從而在行為層面引導用戶選擇^{[10][12]}。使用者對內容的回應因此並非完全自主，而是在演算法與介面共同塑造的結構性行為空間中發生。

從系統演化角度看，強化反饋回路是一種自組織過程。演算法通過局部回饋優化實現全域結構穩定，使用者行為在多輪交互中形成規律性分佈^{[9][10]}。這種動態平衡不依賴外部控制，而是由回饋機制驅動維持，使推薦系統表現出複雜適應系統的特徵。

3.3 用戶偏好收斂與演化趨勢

長期運行的推薦系統往往呈現出使用者偏

好收斂的趨勢。演算法在強化過程中放大高回饋行為模式，使使用者接觸的資訊類型逐漸集中。這一收斂趨勢在多個研究中被觀察到，被認為是“回音室”（echo chamber）或“過濾氣泡”（filter bubble）現象的一種演算法性呈現^{[10][13]}。

偏好收斂首先表現為內容多樣性的下降。推薦演算法在優化回饋信號時，會優先選擇產生最大即時回報的內容類別型，而對低回饋內容的機會減少^[14]。高回饋區域獲得更多曝光，而長尾內容或低回饋類型被邊緣化，導致內容聚集效應增強。隨著時間推移，使用者興趣空間不斷壓縮在演算法偏好的核心區間。

偏好收斂的第二層面體現在使用者心理結構的變化。使用者在長期交互中對推薦系統產生信任感與依賴，演算法逐漸成為預設的資訊選擇代理，使用者的探索性逐漸減少^{[9][10]}。這一趨勢與認知科學中的“確認偏誤”效應一致，即人們傾向於接受與自身既有興趣一致的資訊，強化回饋機制加劇了這一心理傾向。

偏好收斂還具有社會擴散效應。當大量使用者在同一演算法邏輯下互動時，個體偏好收斂過程可能在群體層面放大^{[10][14]}。推薦系統通過共用模型參數影響整體使用者行為模式，使群體行為趨同成為全域性現象。平臺生態中出現內容同質化、新穎內容難獲得曝光等現象，反映了強化機制在社會層面的累積效應。

從機制角度看，偏好收斂並非演算法失效，而是強化過程的自然結果。在高回饋區域達到收益最大化時，演算法會壓縮探索空間，使系統在局部最優狀態保持收斂^{[10][14]}。除非引入外部擾動或不同目標函數，否則系統難以自發打破收斂趨勢。這一結構性特徵說明推薦系統的多樣性衰減並非偶然，而是強化機制的內在邏輯。

4 演算法強化與使用者偏好的共演化模型

推薦系統在長期運行過程中逐漸形成一種動態複雜的交互結構。演算法與使用者並非簡單的因果關係，而是在持續的回饋迴圈中相互影響、相互適應、共同演化。演算法通過強化機

制不斷學習使用者的行為模式，使用者在演算法生成的內容環境中改變自身的興趣、注意力與選擇習慣。二者之間的關係既不是單向控制，也不是完全獨立，而是一種持續耦合的動態系統。這種現象可以用“共演化”來描述^{[9][10]}。

4.1 共演化的概念界定

“共演化”一詞最早源於生物生態學，指不同物種在競爭、合作等互動中共同演化的過程。在複雜系統科學中，該概念被擴展為兩個或多個子系統在長期交互中通過回饋調整模式與狀態，實現相互塑造和動態演化^[9]。在推薦系統語境下，這種動態過程體現為演算法與使用者之間的交互回饋迴圈。

在推薦系統中，演算法根據行為資料生成推薦結果，使用者對推薦內容的選擇、點擊、停留和忽略等行為又成為新一輪訓練資料。這些信號推動模型參數更新，導致系統輸出狀態發生變化，從而影響用戶下一輪的行為環境。隨著這一過程的反覆運算，演算法與使用者的狀態不斷調整，形成長期自我調整互動^[10]。演算法在優化過程中不斷調整參數，而使用者在資訊暴露與心理反應中不斷重塑興趣結構，二者共同構成一個自組織的演化系統^[9]。

共演化的關鍵特徵是相互依賴與回饋放大。演算法需要使用者行為作為學習素材，而使用者的行為模式又在演算法輸出下發生變化，這種迴圈體現出非靜態平衡的動態調整特性^[9]。演算法不是簡單地適應使用者，使用者也不是被動接受演算法，而是在多輪交互中共同決定系統演化方向。共演化體現了人機之間的共構關係，演算法強化與人類認知行為在同一回饋網路中進行共同演進。

在這一結構中，推薦系統不再是外部為使用者服務的工具，而成為人類行為的一部分。使用者興趣不再是獨立生成的心理狀態，而是演算法結構中不斷被重塑的變數^[10]。演算法的學習規則與使用者行為策略交織成一個複雜的共演過程，系統穩定性、偏好分佈以及資訊生態特徵都源自這一共演化機制。

共演化的另一核心維度是自我調整性。演算法與使用者在互動過程中不斷調整自身行為規則以適應對方變化。演算法通過強化學習或

深度學習機制更新推薦策略，而用戶在推薦流中形成對演算法輸出的心理預期，並據此調整自身反應。系統內部每一方通過經驗積累實現適應性演化。演算法優化速度與使用者行為變化頻率共同決定系統動態平衡狀態，當兩者適應節奏同步時，系統趨向穩定；當變化速度不匹配時，系統可能出現震盪或模式漂移。

共演化不僅是一種關係描述，也是一種生成機制。它揭示了推薦系統如何通過長期交互形成複雜行為結構。演算法強化的技術邏輯與使用者的心理反應共同構成系統演化動力，理解共演化核心在於識別回饋路徑與適應過程之間的非線性關聯^[10]。

4.2 人 - 演算法系統的動態耦合機制

推薦系統中的人 - 演算法關係可以被視作一種動態耦合結構。演算法是通過數學模型實現的自學習系統，使用者是具有情感、認知與動機的行為體。兩者的互動不是線性的輸入 - 輸出關係，而是一種相互影響、相互重塑的動態過程^[12]。演算法的更新規則決定了使用者在平臺中的可行行為空間，用戶行為變化又反過來影響演算法參數，使模型對特定特徵的回應敏感性不斷調整。

這種動態耦合體現於多個層面。第一層是結構耦合。推薦系統通過介面與交互設計將演算法輸出嵌入使用者體驗中，用戶在操作過程中無意識適應系統邏輯。演算法的內容佈局、推薦節奏與呈現方式構成行為場景，使用者反應被收集為資料回饋^[10]。

第二層是認知耦合。使用者的注意力、興趣與判斷在演算法推薦引導下形成新的感知模式。演算法決定資訊呈現順序，從而塑造使用者心理預期。使用者選擇行為並非完全自主，而是在演算法生成的注意力環境中發生。演算法持續調整推薦權重，使使用者認知傾向逐漸與系統輸出相一致。使用者通過習慣與學習形成對演算法的適應性反應，演算法也在使用者聚合行為中學習回饋規律。這種認知層面雙向適應使系統呈現複雜的心理動力特徵。

第三層是演化耦合。演算法強化過程中通過優化目標函數更新模型參數，使用者在內容消費中形成新的行為模式。隨著交互累積，演

算法與使用者狀態變數在高維空間形成相互依賴軌跡。演算法強化速度與使用者行為變化率共同決定系統演化方向。當演算法過度敏感時，系統可能陷入短期回饋迴圈；當使用者行為變化太慢時，演算法可能失去適應性，導致推薦品質下降^[12]。系統演化路徑因此呈現多重非線性動態。

動態耦合機制使推薦系統具備自組織特徵。系統整體狀態不是外部控制的結果，而是演算法與使用者相互作用的產物。微小演算法調整可能引發大規模行為偏移，個體用戶反應聚合又可能反向影響模型參數分佈^{[9][12]}。這種非線性互動構成系統複雜性根源，長期演化過程難以完全預測。

從資訊理論視角看，動態耦合可以視為資訊熵的雙向傳遞過程。演算法從使用者行為中獲取資訊以降低預測不確定性；使用者從演算法推薦中獲取內容以提升認知確定性。隨著回饋迴圈持續，系統總體熵值趨向穩定，表現為推薦準確度提升與用戶行為慣性化^[10]。系統熵平衡狀態代表共演化階段性穩定，但這種穩定是暫時的，任何新的演算法策略或使用者群體變化都可能打破平衡，引發系統重構。

耦合機制還具有社會性特徵。使用者並非孤立個體，而是處在社交與行為網路中。演算法對個體行為的回應通過群體聚合效應影響整體行為分佈，群體行為變化又反向調節演算法更新。這種社會層面耦合使推薦系統演化具備集體動力^[14]。演算法在優化個體體驗的同時，也在無意中調整群體資訊流結構，使共演化不僅發生於個人 - 系統層面，也在技術系統與社會系統之間展開。

4.3 共演化機制的工程特徵

從工程角度看，演算法與使用者的共演化過程表現出三個顯著特徵：自我調整性、非線性與難以逆轉性，這些特徵使推薦系統的演化不同于傳統線性優化結構，而呈現複雜自組織系統模式^{[9][10]}。

自我調整性表現為系統根據環境與使用者行為變化調整內部狀態的能力。演算法強化學習中不斷調整策略優化長期回饋，用戶在推薦流中通過習慣形成、興趣更新或選擇遷移來適

應系統輸出。雙向調整形成動態平衡，使推薦系統能夠在不斷變化的內容與需求環境中保持有效性。

非線性體現為系統對微小輸入變化的放大反應。一方面，演算法的微小參數調整可能引起大規模行為偏移；另一方面，使用者群體行為的變化通過資料回饋誘發模型結構改變^{[12][14]}。這種非線性源於回饋機制的多層嵌套，即系統不僅回應個體使用者行為，還回應群體資料分佈。延遲回饋進一步加劇動態反應滯後性。

難以逆轉性表現為系統狀態的路徑依賴。一旦演算法強化形成特定參數結構或偏好模式，系統難以通過局部調整回到初始狀態；長期交互中使用者形成的行為慣性使系統對規則變化的回應具有遲鈍性^[14]。這種路徑依賴使推薦系統在長期演化中形成穩定但封閉的狀態空間。

共演化機制具有分層結構性。在個體層面，演算法與單個使用者之間回饋構成微觀共演；在群體層面，使用者行為聚合與演算法全域參數調整構成宏觀共演。局部行為模式與全域系統結構之間存在跨層回饋，個體行為影響演算法方向，而演算法更新又通過策略影響個體行為。

工程意義上，這種共演化機制解釋了推薦系統的自組織與複雜性來源。演算法強化不僅是優化工具，更是系統內部演化動力。使用者偏好不是靜態變數，而是在演算法學習過程中不斷重構的動態要素。推薦系統的穩定性、偏好收斂與多樣性衰減都可以從共演化視角理解。演算法與使用者共同構成具有自我調節能力的適應系統，其行為由內部機制驅動而非外部干預。

這種共演化機制在大規模平臺上尤為顯著。當使用者數量與內容規模超過一定閾值時，系統的回饋網路呈現高維複雜性。演算法優化的每一次反覆運算都在重塑資訊流結構，使用者行為聚合反過來驅動演算法演化。長期運行中形成宏觀模式，如內容集中化、興趣分層與注意力極化^[14]。

從長期演化視角看，共演化機制意味著推薦系統處於持續未完成狀態。演算法與使用者

的互動持續生成新平衡點，在穩定與擾動之間迴圈。任何外部干預如演算法規則調整或介面設計變動都可能引發新的演化方向。推薦系統因此具有演化開放性，其結構在不斷自我調整中維持生命力^{[9][10]}。

理解演算法強化與使用者偏好的共演化機制，有助於從根本上認識推薦系統的長期動態。本章揭示了技術系統與人類行為之間的深層耦合關係，說明推薦演算法不只是資訊工具，而是認知環境的一部分；演算法強化塑造了使用者偏好結構，而使用者偏好又通過行為回饋塑造演算法邏輯，共演化過程在雙向塑造中持續展開。

5 機制分析的工程啟示與系統思考

推薦系統的長期運行結果往往超出了工程設計的初始預期。演算法在不斷強化的過程中不僅優化了性能指標，也改變了系統的整體演化方向。機制分析揭示出，推薦系統的行為是目標函數、資料結構、回饋信號與使用者反應之間的複雜互動產物。理解這一互動機制的工程啟示有助於從系統層面反思演算法設計的邊界與潛在後果。推薦系統不再是單純追求效率的計算模型，而是一種具有演化屬性的動態系統，其行為應當從長期結構和演化趨勢角度進行分析。

5.1 強化目標與系統行為之間的關係

推薦系統的長期運行結果往往超出了工程設計的初始預期。演算法在不斷強化的過程中不僅優化了性能指標，也改變了系統的整體演化方向。機制分析揭示出，推薦系統的行為是目標函數、資料結構、回饋信號與使用者反應之間複雜互動的產物。理解這一互動機制的工程啟示，有助於從系統層面反思演算法設計的邊界與潛在後果。推薦系統不再是單純追求效率的計算模型，而是一種具有演化屬性的動態系統，其行為應從長期結構和演化趨勢角度進行分析^[6]。

目標函數與系統行為之間並非線性映射，而是複雜的回饋系統。演算法的優化過程反向影響使用者行為模式，使用者行為再通過回饋信號影響模型更新。目標函數的定義實際上決

定了演算法對環境的敏感區域，也決定了系統的長期演化方向。例如，當推薦系統目標是最大化點擊率時，演算法傾向推送高吸引力內容，使用者逐漸適應這種刺激結構，形成短時偏好，系統整體逐漸偏離對真實興趣的刻畫。

目標函數的設計不僅影響系統性能，也影響其社會與心理效應。點擊率或停留時間等指標雖易量化，卻無法代表用戶長期滿意度或認知收益。演算法通過強化機制優化這些指標，實質上是在優化用戶即時反應，而非資訊品質。隨著時間推移，系統在高回饋區域內收斂，形成封閉的偏好迴圈。使用者行為在演算法引導下逐漸同質化，系統失去了對多樣化興趣的敏感性。

目標函數的單一化是強化機制偏移的根源。當系統忽視多目標平衡時，演算法行為方向被短期信號主導，表現出路徑依賴性。模型參數在特定優化區域內不斷強化，形成長期結構偏移。工程實踐表明，目標函數的微小調整往往引發系統級行為變動，這種非線性關係說明演算法優化與系統行為之間存在深層耦合。

一個更系統的工程視角應將目標函數視為系統演化的控制變數。其設計需考慮強化機制的動態特徵，包括回饋延遲、使用者慣性和系統自組織趨勢。演算法優化的真正目標應從性能提升轉向系統穩定性與長期演化方向的調控。

5.2 推薦系統穩定性與多樣性問題的機制解釋

推薦系統在長期運行中普遍出現多樣性下降、內容集中和興趣趨同等問題。這一現象表面上似乎是演算法偏差或資料不均衡造成的，但從機制角度看，是強化回饋自然演化的結果。演算法在強化學習過程中聚焦高回報區域，模型的注意力被限制在回饋最強部分，系統因而逐漸喪失探索能力。

推薦系統的核心矛盾在於探索與利用的平衡。演算法需在已知高回饋內容與潛在未知內容之間分配資源。強化學習傾向於利用歷史經驗，減少不確定性，而這一傾向在長期中會抑制探索性。系統在短期收益驅動下形成自我驗證的偏好回路，探索空間被壓縮，多樣性下降並非設計缺陷，而是回饋回路內生結果。

系統穩定性與多樣性之間存在天然張力。強化機制通過優化回饋信號實現穩定收斂，但穩定往往以犧牲多樣性為代價。收斂速度越快，內容覆蓋面越窄，使用者行為模式越單一。演算法在追求性能中形成穩態結構，這種穩態雖提升了預測精度，卻削弱了系統的動態適應性。當外部環境或興趣發生變化時，系統調整遲滯、回應慣性增強。

多樣性問題不僅存在於內容層面，也反映在行為結構中。長期交互使用戶對推薦結果的依賴增強，探索意願下降。演算法強化使用者對特定類型內容的偏好，使用者行為反過來驗證演算法判斷，系統陷入偏好收斂的迴圈。多樣性下降因此是演算法與使用者共同作用的產物，系統在回饋放大中自發生成行為同質化結構。

工程上，穩定性與多樣性的平衡應在模型設計階段考慮。單一優化目標易導致系統收斂至局部最優。引入多目標優化、探索性策略或熵正則化機制可緩解問題。系統需在局部回饋最大化與全域結構平衡間維持動態協調，從而保持長期健康性與多樣性。

推薦系統的穩定性還受回饋時延影響。行為回饋具有時間滯後，模型更新使用的信號往往落後於真實興趣變化。短期內演算法可能過度響應舊資料，導致學習振盪。理解並建模回饋時延，有助於提高演算法自我調整能力，使系統在不犧牲多樣性的前提下維持穩定運行。

5.3 對推薦系統設計的機制層面啟示

從機制角度看，推薦系統的設計應當超越性能導向的短期思維，轉向對系統演化規律的整體理解。演算法強化並非僅僅是模型優化的過程，而是系統在使用者行為環境中不斷調整自身結構的演化機制。強化過程本身應被視為關鍵變數，其設計需要考慮對系統長期行為的塑形作用。

工程設計的第一層啟示是目標多樣化的重要性。推薦系統不應僅以點擊率或停留時間作為優化目標，而應引入反映用戶長期滿意度與多樣性維持的指標。多目標優化可以使演算法在短期收益與長期結構之間建立平衡。系統設計者需要明確不同目標之間的權重關係，使演

算法能夠在強化學習過程中保留探索能力。

第二層啟示是回饋結構的可控性。推薦系統的回饋機制決定了演算法與使用者之間的資訊流動路徑。系統設計應避免形成完全閉合的強化回路，需要通過外部擾動或引入隨機探索信號打破自我驗證結構。例如可以通過引入冷開機機制或動態重置策略，讓演算法定期重新採樣使用者興趣空間，從而恢復多樣性與適應性。

第三層啟示是系統時間尺度的設計。推薦演算法的學習週期、更新頻率與使用者回饋節奏之間需要建立協調關係。過度敏感的演算法容易陷入短期波動，過於遲鈍的演算法又難以回應環境變化。合理的時間尺度設計可以防止系統陷入非平衡狀態，使強化過程在長期內保持可控的演化路徑。

第四層啟示是結構透明性。隨著推薦系統的複雜化，演算法內部的強化路徑和行為邏輯逐漸難以解釋。建立可解釋的模型結構有助於工程師識別強化機制的偏移風險，從而在早期階段進行調整。系統的透明性不僅提高工程可控性，也有助於在社會層面增強用戶信任。

推薦系統的設計應當以系統演化的可持續性為核心。演算法強化機制在優化局部性能的同時可能削弱系統的長期生命力。工程思維需要從靜態的性能追求轉向動態的系統調控，強調對演化趨勢的監測與調節。推薦系統的設計不應只關注模型在當前環境下的最優性，而應考慮其在未來環境中的適應性。

機制分析提供了一種新的工程視角。它將推薦系統視為複雜的演化體，而非孤立的演算法集合。理解演算法強化與使用者偏好共演化的機制，有助於構建更具穩態性與多樣性的系統結構。推薦系統的設計應當在優化與演化之間尋求平衡，使技術發展與用戶行為共處於可持續的動態平衡之中。

6 結論

推薦系統的演化已經超越了傳統的演算法優化范式，成為一種人機共生的動態系統。演算法的強化學習機制在長期運行中不只是提升了預測精度和個性化水準，也在不斷重塑使用

者的興趣結構與行為模式。本文通過對演算法強化與使用者偏好互動機制的分析，揭示了推薦系統在運行過程中所呈現出的共演化特徵。演算法與使用者在回饋迴圈中相互依賴、相互塑造，共同構成了系統演化的核心動力。

本文首先從系統結構層面對推薦系統進行了分析。推薦系統由使用者、內容、資料與模型構成的閉環結構，使其具備自學習特徵。演算法以使用者行為為輸入，通過回饋信號實現自我強化。這一過程的本質是一種持續的交互學習。用戶的每一次行為既是個體偏好的表達，也成為演算法更新的依據。系統通過對回饋信號的回應不斷調整推薦策略，使用戶體驗趨向個性化，而使用者的行為又因演算法輸出而發生變化。

機制分析表明，演算法強化與使用者偏好之間存在結構性的共振關係。演算法的優化目標決定了系統的學習方向，而用戶行為在回饋中起到調節作用。系統在高回饋區域不斷強化，形成了興趣收斂的趨勢。這種趨勢不僅表現為內容多樣性的下降，也體現在使用者心理結構的變化。使用者逐漸在演算法生成的內容環境中形成新的興趣模式，探索性行為被削弱，注意力資源集中於演算法偏好的區域。演算法與使用者在多輪交互中形成穩定的回饋路徑，構成共演化的動力結構。

共演化機制使推薦系統呈現出非線性、自我調整與難以逆轉的特徵。演算法的每一次更新都在改變用戶的認知環境，而用戶行為的微小變化又可能引起演算法參數的系統性偏移。系統在這種迴圈中實現自組織和平衡，但這種平衡往往是局部的、暫時的。當外部條件或目標函數發生改變時，系統會重新尋找新的平衡點。推薦系統因此不是靜態優化器，而是不斷重塑自身與使用者關係的演化系統。

從工程視角來看，理解共演化機制的意義在於重新定義演算法優化的目標。推薦系統的長期行為不應被簡單視為性能指標的延伸，而應被理解為演算法與使用者共同適應的結果。單一的目標函數往往導致系統演化方向過於集中，使使用者偏好結構固化。多目標優化、動態回饋調節與探索機制的引入，有助於打破這種收斂趨勢，使系統在長期運行中保持多樣性與

穩定性。演算法的強化過程應被納入系統設計的整體框架中，作為影響系統演化方向的核心變數。

本文的理論貢獻在於提出一種機制性解釋框架，用以理解推薦系統的長期行為特徵。該框架揭示了推薦系統中的優化過程與人類行為之間的深層耦合關係，說明演算法的技術演化與使用者的心理演化是同步發生的。演算法通過對行為資料的強化學習生成推薦模式，使用者在這一模式中形成興趣偏好，系統的長期輸出即是二者共演化的結果。這一視角有助於突破傳統以演算法性能為中心的研究範式，使推薦系統的研究回歸到系統性與動態性的層面。

未來研究可以在此基礎上展開更細緻的建模與實證工作。一方面，可以從系統動力學角

度建立演算法強化與使用者偏好共演化的數理模型，以刻畫回饋放大、偏好收斂與多樣性演化的規律。另一方面，可以通過長期行為資料驗證共演化過程的動態路徑，為演算法調控提供經驗支援。共演化機制的深入理解將為推薦系統的可持續發展提供理論基礎，也為未來人機協同系統的設計提供新的思路。

推薦系統的演化並非線性增長的技術進步，而是一種複雜的社會技術過程。演算法強化機制在優化資訊流的同時，也在重塑認知空間。人類的興趣、選擇與價值感知在演算法回饋中不斷被重新定義。理解這一過程，有助於認識技術系統與人類行為之間的共生關係，也為資訊社會的長期結構性演變提供了新的理論視角。

參考文獻：

- [1] Adomavicius G, Tuzhilin A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6): 734-749.
- [2] Mansoury M, Abdollahpouri H, Pechenizkiy M, Mobasher B, Burke R. Feedback Loop and Bias Amplification in Recommender Systems [EB/OL]. arXiv preprint arXiv:2007.13019, 2020. Available at: <https://arxiv.org/abs/2007.13019>
- [3] 曹俊偉, 李軍祥, 李玉璐. 電子商務推薦系統研究綜述[J]. 電子商務評論, 2025, 14(12): 3087-3097.
- [4] 胡琪, 朱定局, 吳惠蕓, 巫麗紅. 智慧推薦系統研究綜述[J]. 電腦系統應用, 2022, 31(4): 47-58.
- [5] Jiang R, Chiappa S, Lattimore T, György A, Kohli P. Degenerate Feedback Loops in Recommender Systems[EB/OL]. arXiv preprint arXiv:1902.10730, 2019. Available at: <https://arxiv.org/abs/1902.10730>
- [6] 張浪朗. 基於多目標優化及偏好策略的推薦演算法研究[D]. 東華大學, 2022.
- [7] 周銘. 智慧資訊系統中的強化學習演算法在推薦系統中的應用[J]. 資訊系統工程, 2024, (08): 52-55.
- [8] Zhang S, Yao L, Sun A, Tay Y. Deep Learning Based Recommender System: A Survey and New Perspectives[J]. ACM Computing Surveys, 2019, 52(1): 1-38.
- [9] Pedreschi D, et al. Human-AI coevolution: Feedback loops and societal impact [J]. Artificial Intelligence, 2025, 339: 104244.
- [10] Krauth K, Wang Y, Jordan M I. Breaking Feedback Loops in Recommender Systems with Causal Inference [EB/OL]. arXiv preprint, 2022. Available at: <https://arxiv.org/abs/2207.01616>
- [11] Mollabagher A, Naghizadeh P. The Feedback Loop Between Recommendation Systems and Reactive Users [EB/OL]. arXiv preprint, 2025. Available at: <https://arxiv.org/abs/2504.07105>
- [12] Barlacchi G, Lalli M, Ferragina E, Giannotti F, Pappalardo L. A Simulation Framework for Studying Systemic Effects of Feedback Loops in Recommender Systems [EB/OL]. arXiv preprint, 2025. Available at: <https://arxiv.org/html/2510.14857v1>

- [13] Jiang R. Degenerate Feedback Loops in Recommender Systems [C]. AAAI/ACM Conference on AI, Ethics, and Society, 2019. Available at: https://www.aies-conference.com/2019/wp-content/papers/main/AIES-19_paper_187.pdf
- [14] Klimashevskaya, A., Jannach, D., Elahi, M. et al. A survey on popularity bias in recommender systems [J]. User Model User-Adap Inter, 2024, 34, 1777-1834.

版權聲明

© 2025 作者版權所有。本文依據“知識共用署名 4.0 國際授權合約”（CC BY 4.0）以開放獲取方式發佈。該許可允許使用者在任何媒介中自由使用、複製、傳播與改編文章（含商業用途），惟須明確署名原作者及出處，並注明所作修改（如有）。完整協議詳見：<https://creativecommons.org/licenses/by/4.0/deed.zh-hans>

出版聲明

所有出版物中的陳述、觀點及資料僅代表作者及供稿者個人立場，與 Brilliance Publishing Limited 及/或編輯人員無關。Brilliance Publishing Limited 及/或編輯人員對因內容所提及的任何理念、方法、說明或產品所導致的人身或財產損害概不負責。

DOI: 10.53104/xdkxsts.2025.01.02.005

國產 AI 晶片上的圖優化編譯技術研究——以昇騰 CANN 為例

周映竹¹

1. 大連理工大學，遼寧 大連，116024

摘要：國產 AI 晶片的規模越來越大，模型結構也越來越複雜。編譯系統在晶片性能和能效優化中起著重要作用。國產 AI 晶片在硬體設計上已經有了明顯的進步，但性能提升仍受到編譯優化的限制。本文以昇騰 CANN 編譯體系為例，研究圖優化編譯在能效提升和軟硬體配合中的原理和方法。

研究從計算圖結構和編譯理論出發，分析圖優化在運算元融合、圖結構調整和記憶體管理中的作用。圖優化能在演算法和硬體之間建立一種對應關係，使計算更高效。通過分析昇騰 CANN 的體系結構，可以看到它的優化重點是以圖為核心的資源配置策略。它通過運算元融合和資料流程調整，讓執行速度更快，能耗更低。

研究構建了一個多目標約束的能效優化模型，把計算代價、頻寬使用和功率消耗放在同一個分析框架中。這樣可以讓圖優化過程更有規律，也能更好地平衡性能和能耗。

在比較研究中，本文對 CANN、TensorRT、TVM 和 NeuWare 等系統進行了分析。結果顯示，國產編譯系統更注重硬體特徵和能效優化，而國外框架更重視通用性和跨平臺相容性。這種不同說明國產編譯系統處於以性能為主要目標的階段，並在逐步提升開放性。

研究還提出了未來圖優化的發展方向，包括自動優化機制、跨晶片統一編譯框架，以及以能效為重點的自我調整編譯系統。圖優化編譯是提升 AI 晶片算力利用率的重要環節，也是推動國產 AI 算力自主化的關鍵技術。通過更完善和智慧的編譯系統建設，可以在能效、性能和生態相容性之間實現新的平衡。

關鍵字：昇騰 CANN；圖優化編譯；能效建模；運算元融合；NPU 架構；國產 AI 晶片

Research on Graph Optimization Compilation Techniques for Domestic AI Chips: A Case Study of Ascend CANN

ZHOU Ying-zhu¹

1. Dalian University of Technology, Dalian 116024, P.R.China

Correspondence to: ZHOU Ying-zhu; Email: zzyzhu0112@163.com

Abstract: The scale of domestic AI chips is growing rapidly, and their model structures are becoming increasingly complex. The compiler system plays an important role in improving chip performance and energy efficiency. Although domestic AI chips have made significant progress in hardware design, their performance improvement is still limited by compiler optimization. This paper takes the Ascend CANN compiler system as an example to study the principles and methods of graph optimization compilation in energy efficiency

收稿日期：2025-12-08 返修日期：2025-12-23 錄用日期：2026-01-08 出版日期：2026-01-14

通信作者：zzyzhu0112@163.com

引用格式：周映竹. 國產 AI 晶片上的圖優化編譯技術研究——以昇騰 CANN 為例[J]. 現代科學探索, 2025, 1(2): 49-61.

improvement and software-hardware coordination.

The research starts from the structure of computational graphs and compilation theory, analyzing the role of graph optimization in operator fusion, graph structure adjustment, and memory management. Graph optimization can establish a correspondence between algorithms and hardware to make computation more efficient. By analyzing the architecture of Ascend CANN, it is found that its optimization focuses on a graph-centered resource allocation strategy. Through operator fusion and data flow adjustment, it achieves higher execution speed and lower energy consumption.

A multi-objective constrained energy efficiency optimization model is constructed in this study, integrating computational cost, bandwidth usage, and power consumption into a unified analytical framework. This allows the graph optimization process to be more structured and better balance performance and energy efficiency.

In the comparative study, this paper analyzes systems such as CANN, TensorRT, TVM, and NeuWare. The results show that domestic compiler systems focus more on hardware characteristics and energy efficiency, while international frameworks emphasize generality and cross-platform compatibility. This difference indicates that domestic compiler systems are currently in a performance-oriented stage and are gradually becoming more open.

The study also proposes future directions for graph optimization, including automated optimization mechanisms, unified cross-chip compilation frameworks, and adaptive compilation systems centered on energy efficiency. Graph optimization compilation is a key step to improving the computing power utilization of AI chips and an essential technology for achieving autonomy in domestic AI computing power. Building a more advanced and intelligent compiler system can help establish a new balance among energy efficiency, performance, and ecosystem compatibility.

Key words: Ascend CANN; graph optimization compilation; energy efficiency modeling; operator fusion; NPU architecture; domestic AI chips

引言

人工智慧模型的規模和複雜度不斷增加，對計算能力的需求增長速度已經超過了硬體發展的速度。模型的參數從百萬、億級擴展到千億級，推理和訓練任務對能效提出了更高要求。AI 晶片的性能不再只取決於硬體算力，還取決於編譯系統對計算資源的管理能力。編譯系統連接演算法和硬體，它的優化水準決定了晶片算力的使用效率和整體能耗。國產 AI 晶片在架構設計上進步明顯，但編譯優化的成熟度還需要提升。

已有的 AI 編譯框架，如 NVIDIA 的 TensorRT、Google 的 XLA 和 TVM，已經在 GPU 和異構計算平臺上形成了完整的優化體系。這些框架通常依賴特定的硬體假設，其優化方法與國產 AI 晶片的架構存在差異。國產晶片常使

用張量加速單元和多級緩存結構，以支援高密度矩陣計算和高頻寬資料傳輸。這種結構與國外晶片不同，使得傳統編譯框架在移植時無法充分發揮硬體潛能。在圖優化階段，運算元融合、記憶體管理和執行順序優化難以匹配國產晶片的執行方式。結果是硬體算力較強，但實際能效不理想。

為了解決這個問題，國產 AI 晶片廠商開始建立適合自身架構的編譯體系。華為昇騰處理器的 CANN (Compute Architecture for Neural Networks) 是一個代表性系統。它採用圖優化驅動的編譯方法，通過運算元融合、圖重構和記憶體複用減少訪存開銷和調度延遲，從而提高 NPU (Neural Processing Unit) 的資源利用率。CANN 的出現標誌著國產 AI 編譯技術從指令級優化轉向圖級協同優化的轉變。目前關於 CANN 圖優化機制的研究仍然較少，公開資料多集中在框架使用和運算元開發上，對其優化原理和

能效機制的分析還不充分。

本文以昇騰 CANN 為研究物件，從編譯結構和優化邏輯兩個角度進行分析，探討圖優化在國產 AI 晶片能效提升中的作用和原理。研究目標包括：分析 CANN 圖優化的理論基礎和執行機制，說明其在計算圖層實現軟硬體配合；建立能效建模框架，從計算圖角度研究性能和能耗的關係；對比國際主流編譯系統，分析國產 AI 編譯框架的結構特點和潛在優勢。

本文採用結構分析、演算法建模和比較研究的方法。結構分析揭示 CANN 編譯流程和圖優化的運行邏輯；演算法建模用於表達圖優化的數學模型和能效約束；比較研究以 TensorRT、TVM 等系統為參考，從優化細微性和硬體匹配角度分析 CANN 的特點。本文關注的重點不是框架性能，而是從編譯理論和系統結構的角度，研究國產 AI 編譯系統的能效機制，即通過圖優化提升硬體性能的方式。

1 圖優化編譯的理論基礎

人工智慧模型的計算通常用計算圖（Computation Graph）表示。計算圖通過節點和邊描述運算元之間的資料關係，體現模型的結構，也為硬體執行提供支援。計算圖是一個有向無環圖 $G=(V,E)$ ，其中節點 V 表示計算運算元，邊 E 表示張量的資料流程。每個節點對應一個函數 $f_i:R^{(n_i)} \rightarrow R^{(m_i)}$ ，邊的連接定義了輸入和輸出的路徑。計算圖的拓撲順序決定執行的依賴關係，所有優化都必須保持運算元功能不變。

在深度學習系統中，計算圖既是資料流程的表達形式，也是編譯系統的重要中間表示（IR）。與傳統編譯器的語法樹不同，計算圖直接反映計算依賴和執行結構，更適合表示大規模並行運算。圖的結構決定模型在硬體上的映射方式。節點數量、邊的密度和圖的深度都會影響運算元調度和記憶體分配。計算圖是模型到機器指令的過渡形式，其結構特徵影響算力的利用效率。

圖優化（Graph Optimization）是在不改變模型功能的前提下，通過結構調整和執行順序改變減少計算量、提升能效的過程。優化可表示

為結構映射 $G \rightarrow G'$ ，其中 G' 與 G 語義一致，但執行代價函數 $C(G') < C(G)$ 。代價函數包含計算量、記憶體訪問次數和通信延遲等指標，用來衡量性能消耗。圖优化的核心是重構計算路徑，通過依賴分析去除冗餘計算，讓計算方式更符合硬體資源分佈。

在 AI 編譯系統中，常見的圖優化方法包括運算元融合、圖結構重排和記憶體管理三種。

運算元融合（Operator Fusion）把多個相鄰運算元合併為一個執行單元，減少中間張量生成和記憶體訪問。例如，將卷積、歸一化和啟動函數結合，可以在一次計算中完成多個操作，減少內核調用次數和調度成本。

圖結構重排（Graph Reordering）通過調整節點的執行順序或改變依賴路徑來提升並行度或改善資料局部性，讓硬體執行更高效。

記憶體優化關注張量的存儲使用和緩存調度。編譯器可以根據資料依賴關係複用記憶體空間，或提前載入關鍵資料，降低訪存延遲。

這些優化依賴計算圖的可分析性和硬體結構的可建模性。CPU 和 GPU 的圖優化多集中在控制流展開和寄存器使用，而 NPU 的圖優化更偏向資料流程分析。國產 AI 晶片普遍採用張量加速單元、多級緩存和高頻寬互聯結構，圖优化的重點從通用運算元調度轉向片上任務分配。GPU 優化追求通用性，NPU 優化更關注硬體適配和能效控制。

從編譯理論角度看，圖優化是深度學習編譯系統的核心部分。它不是單一模組，而是貫穿圖解析、運算元生成和調度的核心邏輯。圖优化的作用是把計算圖轉化為硬體可執行的指令圖，讓模型結構可以在晶片上高效運行。通過圖優化，編譯系統能在演算法層和硬體層之間建立穩定映射，為能效和資源配置提供基礎^[1]。

2 國產 AI 晶片與昇騰 CANN 編譯體系概述

國產 AI 晶片的進步與人工智慧應用的需求擴張幾乎同時出現。自 2017 年起，深度學習

在圖像識別、自然語言處理和大模型訓練等領域得到廣泛應用^[2]。國產晶片廠商逐漸形成以計算架構創新和生態建設為核心的發展方向。昇騰、寒武紀、天數智芯、比特大陸等企業在指令設計、算力調度和能效優化上具有共同特徵。第一，晶片普遍使用張量計算單元（Tensor Engine）作為基礎模組，以矩陣乘法為核心運算，提高深度學習中卷積和全連接操作的計算速度。

第二，採用資料流程驅動的調度結構，通過片上緩存和高頻寬互聯減少訪存延遲，提高流水執行效率。

第三，引入分層能效設計，在計算、存儲和通信三部分分配功耗，使晶片在高負載下保持穩定能效。

表 1 展示了當前主要國產 AI 晶片的架構特徵和編譯系統優化方向。

表 1

晶片平臺	製造商	核心架構	峰 值 算 力 (TFLOPS)	編 譯 系 統	優化特徵
昇騰 910	華為	Da Vinci NPU 架構	320	CANN	圖優化驅動、運算元融合、能效調度
寒 武 紀 MLU370	寒武紀 科技	TPU-like NPU	256	NeuWare	指令融合、範本運算元調度
比 特 大 陸 BM1684	比特大陸	ASIC 專用架 構	196	自 研 框 架	靜態流水線、定制運算元優化
天數智芯 BPU	天 數 智 芯	BPU 異構單 元	120	自 研 框 架	頻寬控制與運算元細微性分配

資料來源：各廠商技術白皮書與公開資料。

表中資料表明，國產 AI 晶片多採用 NPU 結構，重點在於提升資料流程效率和能效管理。編譯系統的優化水準決定晶片計算潛能的發揮。與 GPU 的通用編譯方式不同，NPU 依靠編譯器深入分析硬體特性，實現任務映射和圖級優化。運算元融合、圖重排和記憶體調度成為性能提升的主要手段。編譯系統的完善程度已成為評估國產 AI 晶片生態成熟度的重要標準。

華為昇騰系列晶片的 CANN（Compute Architecture for Neural Networks）體系代表了國產 AI 編譯技術的結構化轉型^[3]。CANN 不是單一的編譯器，而是面向 AI 計算全流程的編譯與運行框架。它通過建立演算法、運算元與硬體之間的統一抽象層，實現跨晶片版本的一致優化與執行性能。

CANN 體系由四個核心部分組成：前端 IR 層、圖優化模組、運算元庫（TBE）和後端調度

引擎。

前端 IR 層負責模型解析與中間表示轉換。主流框架如 TensorFlow、MindSpore、PyTorch 匯出的計算圖會被轉換為 CANN 內部的統一表示結構。該層重點在於保持語義一致與依賴分析，為後續圖優化提供基礎。

圖優化模組是系統的核心部分。該模組通過掃描和重構計算圖，執行運算元融合、資料流程重排和記憶體複用等優化。優化過程包含模式識別、結構重寫和執行計畫生成三個步驟。CANN 採用多級優化策略，局部模式識別用於運算元融合，全域拓撲分析用於圖級調度。層次化設計讓系統能在不同細微性上平衡性能與能效。

運算元庫 TBE 負責運算元範本的定義與代碼生成。每個範本描述運算元的輸入輸出、資料佈局和執行配置。編譯系統根據硬體特性選

擇最合適的實現方式。TBE 支援運算元自動生成和多版本管理，減少人工調優。

後端調度引擎負責將計算圖映射為硬體指令。它根據硬體結構、算力單元分佈和頻寬限制，生成可直接運行的調度圖。該模組與運行時系統配合，實現編譯時優化與運行時調度的閉環。

CANN 的設計思想是以圖優化驅動硬體適配。傳統編譯框架依賴手動調度或靜態範本，而 CANN 通過圖結構分析決定運算元融合、記憶體分配和執行順序。編譯器具備識別硬體特徵的能力，而不是依賴固定規則。例如，在矩陣運算密集的模式中，圖優化模組根據運算元依賴自動調整順序，提高緩存命中率，減少外部訪存能耗。編譯器在生成調度計畫時會動態平衡計算密度和頻寬佔用，使結果在能效與性能間取得較優比例。

CANN 的出現讓國產 AI 編譯體系更系統化。編譯系統不再只追求速度或相容性，而是通過結構化機制提升能效。它連接模型語義、計算結構與硬體執行，為高效計算奠定基礎，也為後續的圖優化研究提供參考框架。

3 昇騰 CANN 的圖優化機制分析

3.1 編譯體系與圖優化流程

昇騰 CANN 的編譯體系以計算圖為主要結構單元，通過分層設計實現從模型描述到硬體執行的完整映射。系統遵循前端解析、中間優化和後端調度的邏輯。主要模組包括圖前端解析（Graph Engine）、運算元構建模組（TBE Kernel）和任務調度系統（Runtime Scheduler）。圖優化模組貫穿整個編譯過程，負責模型結構調整和資源映射。

在前端階段，CANN 通過 Graph Engine 完成模型的語義解析和圖構建。來自 MindSpore、TensorFlow 或 PyTorch 等框架的模型首先被轉換為統一的中間表示（IR）。該階段保證語義一致和結構完整。Graph Engine 對模型進行拓撲排序，建立節點依賴，並完成類型推斷和運算元分類。每個運算元節點被定義為一個計算單元，包含計算邏輯、輸入輸出資訊和執行約束。這一結構為圖優化階段提供了基礎。

中間層的圖優化模組是 CANN 的核心。它的目標不是單個運算元的微調，而是從整體計算圖角度分析依賴並進行資源映射。優化分為三個步驟：模式識別、圖重寫和計畫生成。

在模式識別階段，系統檢測計算圖中的典型結構，如卷積塊、殘差單元或矩陣乘加鏈，識別可以融合或重排的運算元組。

在圖重寫階段，編譯器根據識別結果調整計算路徑，把獨立節點融合為複合運算元或調整執行順序，減少多餘計算和資料搬移。

在計畫生成階段，系統依據硬體資源約束，如 NPU 核心分佈、存儲容量和頻寬，生成執行圖。計算圖此時被劃分為執行塊（Execution Block），形成硬體可識別的調度計畫。

運算元構建階段由 TBE 模組完成。TBE 將優化後的運算元定義轉換為計算內核（Kernel）。每個內核包含計算邏輯、資料佈局和類型設置。CANN 在此實現運算元範本化和自動生成，讓運算元能根據張量大小和精度動態調整。圖優化和運算元構建互相影響。圖優化確定運算元融合的範圍與輸入輸出關係，TBE 根據結果生成相應內核。運算元執行性能會反過來影響圖優化策略，形成結構與性能之間的迴圈回饋。

圖優化與後端調度之間的銜接體現了 CANN 的系統設計。後端調度模組將優化後的執行圖映射到硬體單元，包括計算核心、存儲資源和通信鏈路。CANN 使用基於硬體拓撲的調度演算法，將執行塊分配到不同 NPU 核心上。調度過程考慮資料依賴與資源使用情況，保持並行度與記憶體佔用的平衡。昇騰的調度機制依賴靜態圖分析結果，執行順序在編譯階段確定，避免運行時調度帶來的額外能耗。

圖優化在 CANN 體系中的作用主要體現在三個方面。

第一，圖優化是模型語義和硬體結構之間的連接中心。沒有圖級優化，運算元編譯無法實現跨層協調，硬體資源會被分散使用。

第二，圖優化是能效控制的主要環節。通過減少訪存、優化資料傳輸和提升緩存命中率，編譯系統在圖層即可完成能耗控制。

第三，圖優化為後端調度提供結構約束。

優化後的執行圖明確任務邊界和依賴，使調度器在編譯階段生成高效執行計畫。

3.2 運算元融合與圖級重構機制

在昇騰 CANN 編譯體系中，運算元融合（Operator Fusion）是圖優化階段的主要策略。它的作用是減少記憶體訪問和任務調度開銷。深度學習模型中包含大量相鄰的線性運算元和非線性運算元，如果這些運算元分別執行，會導致計算單元頻繁交換資料。每次運算元切換都會引起張量的中間存儲、緩存刷新和 DMA（Direct Memory Access）操作，造成性能損失。CANN 通過分析運算元依賴和資料流程特徵，把連續的運算元合併為一個執行單元，使資料直接在片上緩存中傳遞，避免多次訪存。

運算元融合分為兩種類型：線性運算元融合和複合運算元融合。線性融合針對順序連接、資料形態相同的運算元組合，例如矩陣乘加（MatMul + Add）或啟動鏈（ReLU + Dropout）。這些運算元在圖中有固定的依賴關係，融合後形成一個運算核，減少張量創建和釋放次數。複合運算元融合常用於卷積網路，將卷積、批歸一化和啟動函數（Conv + BN + ReLU）組合在一起。這樣卷積結果能直接進入啟動操作，不需要寫入外部存儲。CANN 檢測到這些結構後會自動識別融合機會，並在圖優化中進行結構修改。

融合策略受到網路結構和資料特性的影響。在卷積神經網路中，融合的目標是降低頻寬使用；在全連接網路中，重點是提高運算元複用率和緩存命中率。CANN 在不同網路類型下使用不同融合策略。在 ResNet 網路中，系統更傾向於融合短路徑卷積塊，同時保持跨層連接的獨立性，以保證梯度計算的準確性。在 Transformer 模型中，融合集中在矩陣乘和加法操作，以減輕注意力層的記憶體壓力。這種策略體現了 CANN 對模型結構的識別能力，可以根據圖結構自動選擇合適的融合方式，而不是使用固定範本。

運算元融合減少了訪存次數，也為圖級重構提供了結構基礎。融合後的運算元成為計算密集型節點，輸入輸出依賴更少，便於重排。CANN 利用這一特性，在全域分析階段重新安

排依賴路徑。圖重構的主要目標是優化計算順序和資料流程，使任務在硬體上實現最大並行度和更高緩存利用率。

在圖重構中，編譯器建立依賴矩陣，對計算節點進行拓撲排序。系統根據節點的資料量、訪問頻率和依賴深度，動態調整執行順序。例如，當多個運算元共用輸入資料時，系統會讓它們在同一週期執行，以提升緩存命中率。對於依賴路徑較長的計算，編譯器會設置分區點，將任務拆分以減少等待。

CANN 的圖重構結合了靜態優化和硬體限制。NPU 採用流水化執行結構，計算與訪存有嚴格的時間關係。如果計算順序不合理，會造成流水線停頓。圖重構通過分析運算元的執行時間，重新安排資料傳輸和計算次序，使指令和資料傳輸在時間上保持同步。這樣能在不改變模型功能的前提下提高執行效率。

運算元融合與圖重構構成 CANN 圖優化的核心框架。融合減少節點數量和中間資料，圖重構優化執行順序和平衡資源使用。兩種機制共同作用，使計算圖結構更加緊湊，編譯器能夠生成能耗更低、性能更穩定的調度方案。

3.3 記憶體複用與資料流程調度策略

在昇騰 CANN 編譯體系中，記憶體優化和資料流程調度是圖優化的重要部分。NPU 架構不同於 GPU，它的運算元調度更密集，片上存儲限制更嚴格。CANN 在編譯階段通過分析張量的生命週期，結合靜態記憶體規劃和動態複用，建立面向能效的記憶體管理模型。

張量生命週期管理是記憶體複用的基礎。每個計算節點對應多個輸入和輸出張量，它們的存在時間由依賴路徑決定。傳統編譯框架通常採用獨立的分配和釋放策略，每個運算元在執行時單獨申請記憶體，結束後釋放。這種方式實現簡單，但在大型模型中容易產生記憶體碎片並浪費頻寬。CANN 在圖優化階段建立生命週期分析圖（Lifetime Graph），通過遍歷計算圖確定每個張量的起始和結束時刻。系統根據這些時間段計算重疊關係，並把不會同時存在的張量映射到同一物理空間。這樣，張量在執行過程中可以迴圈使用片上存儲，減少記憶體佔用和片外訪問。

靜態配置的優點是執行穩定。編譯階段完成的記憶體映射表能保證執行時不需要額外位址調整。但靜態方法無法處理動態形狀或條件執行結構。CANN 在運行時引入動態複用機制 (Dynamic Memory Reuse)，由運行時記憶體管理器監控張量狀態。當某個張量計算結束且不再被引用時，系統立即回收空間並分配給新的張量物件。這個過程自動完成，不需要人工干預。靜態規劃提供全域框架，動態複用提供靈活調整，兩種方式組合成成分層記憶體優化體系。

在資料流程調度中，CANN 使用拓撲驅動的調度方法。編譯器通過依賴關係確定執行順序，並在生成執行計畫時進行資料流程重排 (Dataflow Reordering)。目標是讓資料在片上盡可能短路徑傳輸，減少 DMA 傳輸次數。系統根據運算元之間的張量複用關係和訪問頻率，對資料流程重新佈線，讓頻繁交互的運算元鄰近執行，減少資料在不同存儲層間的移動。例如，在卷積、批歸一化和啟動函數連續的結構中，CANN 會將三者組合為連續計算區間，使張量只在片上 SRAM 中流動，不進入外部記憶體。

CANN 的資料流程調度依賴硬體拓撲資訊實現軟硬體配合。昇騰 NPU 的內部存儲包括多級緩存和統一控制器，各核心共用片上存儲但頻寬有限。編譯器根據硬體描述檔 (Hardware Descriptor) 建立頻寬模型，計算各階段的訪存負載。資料流程重排演算法以頻寬均衡為目標，在任務間插入延時指令或調整執行順序，避免訪問衝突。這種方法雖然複雜，但能顯著降低能耗和延遲。

CANN 與 GPU 編譯體系存在明顯差別。GPU 使用統一顯存和多級緩存，優化重點是執行緒並行和訪存合併；NPU 採用固定運算元流水線和任務塊調度，記憶體分配必須在編譯階段確定。GPU 傾向於運行時調優，CANN 強調編譯時確定。GPU 通過動態調度平衡負載，CANN 通過靜態圖優化控制能耗和執行時序。兩種架構的核心區別在執行模型：GPU 依靠多執行緒隱藏延遲，NPU 通過資料流程控制能效。CANN 通過生命週期管理、靜態規劃和動態複用結合，實現了 NPU 資源的精確調度。

3.4 圖優化與硬體協同邏輯

昇騰 CANN 的圖優化機制不僅進行編譯層的結構調整，還直接參與硬體執行的資源配置。其目標是讓計算圖與 NPU 架構實現高效映射，使運算元執行與硬體調度緊密配合。編譯器在這一機制中不只是代碼生成工具，而是控制能耗和性能的核心調度系統。

CANN 在運算元調度階段採用分層映射方法。優化後的計算圖被拆分為多個子圖 (Subgraph)，每個子圖對應一組關聯度高的運算元。編譯器將這些子圖分配到不同的 NPU 計算單元中，包括 Cube 單元、Vector 單元和 Scalar 單元。運算元類型、資料形式和並行特性決定計算資源配置。例如，卷積運算元主要在 Cube 陣列上執行，資料預處理和啟動函數由 Vector 單元完成。通過這種分層方式，系統在物理結構上實現平行計算。

在運算元調度和指令發射之間，CANN 使用指令依賴分析 (Instruction Dependency Analysis)。編譯器依據計算圖拓撲生成依賴矩陣，標識運算元間的順序與資料傳遞。調度模組據此確定指令執行次序，使發射順序與計算圖一致。系統通過硬體指令流 (Instruction Stream) 執行，以固定步長逐條發射指令，保證不同計算單元同步運行。這種靜態調度方式消除了運行時動態調度的負擔，使能耗分佈更穩定。

片上緩存結構是圖優化與硬體協同的核心。昇騰 NPU 採用多級緩存架構，包括統一緩存 (UB)、全域記憶體 (GM) 和寄存器組。CANN 通過圖優化確定 Tensor 在各層緩存中的駐留位置，使資料流程動和計算過程保持一致。系統在執行計畫生成時，會根據運算元融合與圖重排結果，計算張量訪問頻率和生命週期。高複用數據放在 UB 中，低複用或臨時數據放在 GM 中。編譯器根據緩存大小和並行度調整資料塊劃分，減少計算單元間的訪存衝突。這種規劃方式讓緩存命中率更高，能耗更低。

Tensor 的搬移在 NPU 執行中消耗較多能量。CANN 通過資料流程追蹤機制識別跨運算元或跨子圖的資料路徑，並利用指令融合與資料預取減少搬移成本。當系統檢測到運算元間存在頻繁資料交互時，會生成複合訪存指令，使多個運算元共用一次資料訪問。同時，CANN

預測下一步任務的資料需求，提前觸發 DMA，將資料從 GM 載入到 UB 中，減少等待時間。Tensor 搬移由此從被動操作變為主動調度，使計算與資料流程保持同步。

CANN 的核心思想是讓“計算圖即硬體調度圖”。圖優化不僅調整運算元依賴關係，還定義硬體執行的時間和空間佈局。每次圖結構變化都會影響指令順序、緩存佈局和頻寬使用。編譯器與硬體之間形成動態平衡關係。能效表現正是這種平衡的結果：當圖優化實現高密度資源佔用與低冗余訪存時，晶片功耗自然下降。

與 GPU 相比，CANN 的協同邏輯更具確定性。GPU 依賴運行時執行緒調度和記憶體合併，由硬體自動分配資源。CANN 在編譯階段完成任務劃分和資料路徑規劃。GPU 強調通用性，CANN 強調結構控制與能效優化。GPU 通過大規模執行緒換取性能，NPU 通過結構優化提升能效。CANN 將演算法邏輯嵌入硬體調度過程，使能效控制成為編譯系統的一部分。

4 圖優化的模型化與能效邏輯

4.1 計算圖優化的形式化描述

在編譯系統中，計算圖(Computation Graph)用於表示深度學習模型的結構和依賴關係。它由節點(Node)和邊(Edge)組成。節點表示運算元或張量操作，邊表示資料依賴或控制關係。對 CANN 來說，圖优化的主要任務是在保持模型語義不變的前提下，對圖進行結構調整，使執行代價在硬體限制下最小化。

設計算圖 $G=(V,E)$ ，其中 $V=\{v_1, v_2, \dots, v_n\}$ 為運算元集合， $E \subseteq V \times V$ 表示資料依賴集合。每個節點 v_i 具有屬性 $\phi(v_i)=\{t_i, c_i, m_i\}$ ，分別表示運算元類型、計算代價和記憶體佔用。邊 $e_{ij}=(v_i, v_j)$ 具有權重 w_{ij} ，表示資料傳輸的開銷或頻寬消耗。

圖優化問題可以表示為：

$$(\min) C(G') = \alpha \sum_{v_i \in V'} c_i + \beta \sum_{e_{ij} \in E'} w_{ij}$$

$$\text{"s.t." } \Psi(G') \equiv \Psi(G)$$

其中 G' 為優化後的計算圖， $\Psi(G)$ 表示模型語義結構。約束條件保證優化後圖與原始

圖功能一致。目標函數中 α 和 β 為權重係數，用於平衡計算與訪存代價。

這一形式表明，CANN 的圖優化是一個受語義約束的多目標優化問題。變數包括運算元順序、融合方式、張量佈局和資料流程方向。

CANN 的優化可分為三個部分：

(1) 圖結構變換(Graph Transformation)：通過模式匹配識別可融合運算元集合 $\Omega = \{S_k \subseteq V \mid f(S_k) \rightarrow v_k'\}$ ，並生成新的節點 v_k' 。

(2) 記憶體複用規劃(Memory Reuse Planning)：為每個張量 T_i 分配物理空間 $p(T_i)$ ，當且僅當生命週期區間 $l(T_i)$ 與 $l(T_j)$ 不重疊時，滿足 $p(T_i)=p(T_j)$ 。

(3) 執行順序優化(Execution Ordering)：在保持依賴關係的前提下，尋找拓撲排序 $\pi(V)$ ，使總代價 $C(G')$ 最小。

這種建模揭示了圖优化的層次結構。上層進行語義等價變換，保證邏輯正確；下層負責資源映射與執行調度，實現物理優化。CANN 在編譯階段通過中間展示層(IR)保持這種分離，讓邏輯優化與硬體約束可以獨立求解。

在具體實現中，CANN 的優化器並不直接求解完整目標函數，而採用啟發式搜索和代價傳播(Cost Propagation)結合的策略。系統先通過圖模式匹配確定候選融合方案，再根據硬體描述檔計算頻寬、緩存和算力參數，估計每種融合的收益，最後選擇代價最低的路徑。這個過程可視為一種動態規劃模型^[4]：

$$F(v_j) = (\min)_{\{v_i \in \text{pred}(v_j)\}} \{F(v_i) + \Delta C(v_i, v_j)\}$$

其中 $\Delta C(v_i, v_j)$ 表示節點 v_i 到 v_j 的綜合代價，包括計算、訪存和調度延遲。系統通過代價傳播機制，在全域範圍構建最優執行路徑。

形式化建模讓 CANN 的圖優化具備可量化的分析基礎。計算圖不只是模型轉換的結構，而是演算法與硬體交互的數學描述。編譯器的工作可以看作在語義約束下求解能耗最小化問題。這一思路將圖優化從經驗化的規則體系轉化為可計算的優化過程。

4.2 能效优化的約束模型

在硬體系統中，計算圖的優化必須受資源限制約束。NPU 的性能受到算力峰值、緩存容量和頻寬限制的影響。圖優化中的能效問題可以理解為多重約束下的代價最小化過程，目標是在能耗預算範圍內獲得最優計算效率。

設昇騰 NPU 的資源集合為 $R=\{C_{\max}, B_{\max}, M_{\max}\}$ 。其中 $C_{\max}$ 表示計算單元峰值算力， $B_{\max}$ 表示頻寬上限， $M_{\max}$ 表示可用緩存容量。對於優化後的計算圖 $G'=(V', E')$ ，在編譯階段定義三個代價函數：

$$C_{\text{comp}}(G') = \sum (v_i \in V') c_i / C_{\max}$$

$$C_{\text{mem}}(G') = \sum (e_{ij} \in E') w_{ij} / B_{\max}$$

$$C_{\text{stor}}(G') = (\sum (v_i \in V') m_i) / M_{\max}$$

這三個函數分別表示計算利用率、頻寬佔用率和緩存佔用率的歸一化形式。它們構成能效優化的主要約束條件。若以總能量消耗 $E(G')$ 為目標變數，可寫為：

$$\begin{aligned} (\min) \quad & E(G') = \int_0^T (P_{\text{comp}}(t) + P_{\text{mem}}(t)) dt \\ \text{"s.t."} \quad & C_{\text{comp}}(G') \leq 1, C_{\text{mem}}(G') \leq 1, \\ & C_{\text{stor}}(G') \leq 1 \end{aligned}$$

其中 $P_{\text{comp}}(t)$ 和 $P_{\text{mem}}(t)$ 表示計算與訪存時的功率消耗。該模型說明，CANN 編譯器通過運算元融合和結構調整，使計算與訪存功率保持平衡，而不是追求單一的性能峰值。

在該約束框架中，能效優化可看作多目標規劃問題。定義性能指標 $P(G')$ 與能耗指標 $E(G')$ ，目標函數為：

$$\begin{aligned} (\max) \quad & L(G') = \lambda_1 (P(G')/P_{\max}) - \lambda_2 (E(G')/E_{\max}) \\ \text{"s.t."} \quad & R(G') \leq R_{\max} \end{aligned}$$

λ_1 與 λ_2 是權重係數，用於控制性能和能耗的優先順序。CANN 編譯器不會顯式求解該函數，而是通過啟發式代價估計進行間接優化。當運算元融合降低訪存成本時，系統會調整執行批次以控制頻寬；當緩存佔用過高時，系統通過局部圖拆分分配任務，避免功率峰值。該機制是一種基於回饋的多目標優化過程。

能效優化中，關鍵指標是邊際能效收益 $\eta_i = (\Delta P_i) / (\Delta E_i)$ 。 η_i 表示運算元 v_i 的性能提升與能耗變化比值。當 $\eta_i > 1$ 時，優化方

案帶來正收益；當 $\eta_i < 1$ 時，系統放棄該優化。CANN 在圖優化階段通過圖級代價函數即時更新 η_i ，用於動態判斷優化策略的有效性。通過持續評估 η_i ，編譯器能維持性能與能耗的平衡。

CANN 的約束模型與 GPU 系統不同。GPU 依賴運行時功率控制（如 DVFS）調節能耗，而 CANN 在編譯階段即確定部分功率分佈。能效優化不再依靠運行時調節，而是由編譯器結構性決策控制。CANN 在中間展示層（IR）中嵌入能耗權重，使圖優化與能量管理保持統一邏輯。編譯器生成的執行圖決定任務順序，也決定能量流向^[5]。

約束模型使圖優化從局部性能調整擴展為系統級能效配置。CANN 的優化過程在多個資源維度中尋找平衡點，這個平衡由硬體條件決定，也受編譯器策略影響。能效因此成為系統設計的一部分，而不是計算結果。

4.3 靜態優化與動態優化的比較邏輯

在 AI 編譯系統中，計算圖的執行方式主要分為靜態圖（Static Graph）和動態圖（Dynamic Graph）。兩種方式的區別在於調度時機和能效管理方法。靜態優化在編譯階段完成所有圖結構優化、運算元融合和記憶體調度，運行時執行路徑固定。動態圖在運行時根據輸入或環境變化動態生成計算圖，可即時調整執行計畫。

在能效管理上，兩種方式表現出明顯差異。靜態圖優化的能耗模型在編譯階段可完全展開。編譯器根據全域拓撲結構計算每個運算元的訪存量、算力需求和張量生命週期，並在編譯階段完成能量分配。靜態優化的優勢包括：

- （1）能量預算和算力調度在執行前已確定，硬體在運行初期即可鎖定電壓和頻率，減少波動；
- （2）資料流程路徑固定，緩存複用率高，訪存能耗更低；
- （3）執行順序穩定，可實現高並行而無額外調度開銷。

這些特徵使靜態優化適合昇騰 NPU。該架構以批次處理和張量並行為核心，依賴確定的調度計畫實現高算力密度。

動態圖的能效表現依賴運行時控制。動態圖框架（如 PyTorch、DyGraph）通過即時構圖提供靈活性，但每次圖生成都帶來額外的調度和資源消耗。運行時的記憶體分配、運算元編譯和依賴解析佔用算力，使功率曲線波動。對於高吞吐 NPU，這種不確定性導致能耗波動增大，功率規劃難以保持穩定。

CANN 採用靜態優化路徑，有明確的技術原因。昇騰架構使用多級緩存和片上匯流排，資源劃分固定。靜態圖能在編譯階段確定運算元順序和資料佈局，使片上資源使用最優。CANN 以運算元範本（TBE）和固定圖結構為基礎，通過離線優化生成執行計畫，提升不同模型間的性能複用。靜態圖的確定性還便於功率控制。編譯器在編譯期預測每個階段的功率分佈，系統預設電源與散熱參數，使能效曲線平穩。

靜態優化的局限在於缺乏輸入自我調整能力。不同批量和輸入形狀下，固定圖結構難以充分利用硬體潛力。它也無法在運行中調整優化策略，一旦圖結構鎖定，系統不能依據負載變化重新調度。靜態優化對模型結構變化敏感，輕微修改都會觸發重新編譯，影響部署效率。這些問題在多工推理和流式資料處理中尤為明顯。

一種新的優化方向是自我調整圖優化（Adaptive Graph Optimization）。這種方法結合靜態和動態機制，在保持靜態優化能效的基礎上引入運行時調整能力。其實現方式分為三個層次：

第一，參數化圖結構。在中間展示層（IR）中定義可調節參數，如張量塊大小和並行度閾值，使編譯器在運行時根據硬體回饋調整執行分支。

第二，能效回饋機制。在執行單元中採集功率和延遲資料，將結果回饋給運行時調度模組，用於修正編譯期的代價模型。

第三，增量式優化編譯。保留部分中間優化狀態，實現快速再編譯，而不是完全重新構圖。這樣可在輸入變化時維持高性能。

這種方法為靜態編譯體系引入了動態回應能力。它不改變 CANN 的基本結構，而是在確定性編譯框架中加入能效調節功能，使系統在保持穩定性的同時具備適應性。未來的圖優化系統可能採用分層結構：底層由靜態圖維持硬體穩定，上層由輕量運行時實現策略更新，在能效與靈活性之間取得平衡。

5 國產 AI 編譯系統的比較與反思

國產 AI 編譯系統的設計方向與國際主流框架不同。昇騰 CANN 的核心特徵是以硬體為中心的圖優化邏輯，通過精確建模硬體結構提升性能和能效。國際系統如 TensorRT 和 TVM 注重通用性與跨平臺能力，追求在不同硬體之間維持一致的抽象介面和優化邏輯。這種差異反映了“硬體專用化”和“軟體可攜性”的技術取向。

表 2 展示了 CANN 與主流 AI 編譯系統的差異。

表 2

編 譯 框 架	開發主體	圖優化細微性	融合策略	IR 層次	硬體依賴性	優化目標
CANN	華為	圖級 + 運算元級	自動融合與模式匹配	接近硬體層	強	性能與能效平衡
TensorRT	NVIDIA	運算元級	範本融合	中層	強	推理時延最小化
TVM	Apache	圖級	Pass 重 寫 + AutoTVM	高層	弱	通用性與編譯速度

NeuWare	寒武紀科技	運算元級	固定融合範本	中層	強	硬體利用率最大化
---------	-------	------	--------	----	---	----------

CANN 的優化邏輯貼近硬體語義，耦合度高。編譯器在編譯階段通過靜態規劃確定資料流程路徑與 NPU 指令調度，使硬體資源利用最大化。代價模型由硬體描述檔驅動，反映存儲結構、匯流排拓撲和功率限制。TensorRT 使用啟發式方法，以推理時延最小為目標，主要針對卷積、歸一化、啟動運算元優化。系統在運行時動態選擇 kernel variant 以匹配 GPU 架構。TVM 利用 Relay 中間表示將計算圖分解為可重寫運算式，再結合自動調優實現性能搜索，目標是通用性和編譯速度。NeuWare 採用半靜態圖優化，通過範本運算元描述與運行時重構結合實現性能調節。

CANN 的運算元融合細微性更細，直接受硬體流水線深度和緩存容量影響。編譯器在融合時需滿足資源限制。例如在 Conv-BN-ReLU 鏈中，CANN 根據緩存容量決定是否合併 BN 運算元，以避免訪存衝突。TensorRT 的融合偏向語義匹配，不考慮底層硬體約束。TVM 通過模式匹配實現跨層融合，但受 IR 層級限制。NeuWare 使用預定義範本，編譯效率高但缺乏靈活性。CANN 的策略體現“結構約束下的靈活性”，即在硬體邊界內啟發式選擇局部最優方案。

各系統在中間表示（IR）層次上也存在差異。CANN 的 IR 包含張量維度、訪存拓撲與任務映射資訊，優化器可訪問硬體描述符執行指令級調度。TVM 的 Relay IR 更接近演算法表達層，抽象出高層語義，不直接描述硬體特徵。TensorRT 的 IR 位於中間層，部分結合 GPU 特徵。NeuWare 的 IR 與 CANN 相似，但依賴人工設定檔。

這種差異帶來優化能力和通用性的權衡。IR 越接近硬體，優化精度越高，但跨平臺能力弱。IR 越抽象，通用性強，但難以實現精細調度。CANN 選擇靠近硬體的路徑，是因為當前國產晶片仍處於算力優化階段，優先追求性能和能效。為實現極限能效，犧牲部分通用性是一種工程上的合理選擇。

以硬體為導向的設計也存在問題。CANN 在不同晶片版本間相容性差，硬體升級時需要重新定義運算元範本和優化規則。封閉的生態體系限制外部框架的接入，科研與產業應用之間存在介面缺口。相比之下，TVM 等開源框架通過高層抽象實現快速適配，使新硬體能更快融入生態系統。

未來國產編譯系統可在專用化與通用化之間尋求平衡。一種方向是構建分層編譯架構。上層保留通用 IR 和自動調優機制，下層開放硬體介面以支援針對性優化。開放部分優化介面，讓外部研究者參與演算法優化，可增強生態協作與創新能力。

6 圖優化編譯的趨勢與制度化構想

圖優化編譯技術正在從工程經驗向體系化智慧發展。AI 模型的複雜度提高，硬體架構日益多樣，傳統依賴人工規則和靜態調優的編譯方式難以兼顧算力與能效。未來的編譯系統需要具備自學習、自調節和跨平臺協作能力，使圖優化成為可持續進化的過程，而不是固定規則的堆積。

6.1 自動化圖優化的演化方向

傳統圖優化依靠人工定義運算元融合規則和固定調度範本。性能受限於專家經驗和單一硬體特徵。自動化圖優化成為新的研究重點。未來的圖優化將利用機器學習方法，通過性能資料驅動的回饋機制實現自動調優和自動融合。系統在第一次編譯時採集運算元執行資訊、訪存延遲和能耗資料，建立性能模型。然後，使用強化學習或貝葉斯優化演算法搜索最優融合和調度方案。這種學習式優化方式能在不同晶片和模型間遷移，形成自更新的優化知識庫。

在昇騰 CANN 體系中，這一趨勢已有體現。CANN 部分模組支援基於代價模型的自動融合判斷，但仍依賴人工規則。如果引入基於性能資料的回饋學習機制，編譯器可以在保證語義等價的前提下自動提煉模式，從而在不同 NPU

架構間快速適配。核心挑戰是構建統一的性能度量標準和可泛化的優化表示。只有在性能可對比、優化可復現的條件下，自動化優化才能進入制度化階段。

6.2 跨晶片統一優化框架的構想

國產 AI 晶片生態呈多架構並行格局。昇騰、寒武紀、比特大陸、天數智芯等體系在指令集、記憶體架構和運算元庫上差異明顯，編譯系統各自獨立。模型在不同晶片間遷移時需重新編譯和調優，造成高成本。這種結構性分裂降低了整體效率，也影響生態協同。

建立跨晶片統一優化框架是制度化圖優化的關鍵。核心目標是以中立化中間表示 (IR) 和抽象硬體描述介面為基礎，實現編譯器間的互通與複用。

框架可分為三層：

- 1) **統一中間展示層 (Unified IR)**：抽象計算圖的邏輯結構與資料依賴，不包含具體硬體特徵。
- 2) **硬體描述層 (Hardware Abstraction Layer)**：通過標準介面描述 NPU 資源特性，如指令延遲、緩存容量和功率曲線。
- 3) **自我調整優化層 (Adaptive Optimizer)**：根據 IR 與硬體描述生成優化策略，執行圖重構、運算元融合與能效調度。

這一框架與 CANN 現有閉環優化體系相相容，是國產生態的擴展。若能建立通用介面標準，使不同廠商的編譯器在 IR 和運算元層共用優化邏輯，可顯著提升模型遷移效率，減少重複調優。統一性不是削弱差異，而是通過制度化建立協同基礎，使硬體差異成為優化約束而非障礙。

6.3 面向“能效智慧體”的編譯系統結構

傳統編譯器的任務是將演算法轉化為可執行代碼。隨著 AI 計算能耗成為核心問題，編譯器正在從“性能優化工具”轉變為“能效管理核心”。未來的圖優化編譯系統可形成三層結構：

- 1) **軟硬體協同層**：編譯器與硬體運行時系統建立回饋通道。晶片即時報告功率、溫度

和延遲，編譯器據此調整運算元調度和資料路徑，實現能效閉環控制。

- 2) **能效約束層**：在圖優化階段引入能量預算，將能耗、延遲和算力作為等權目標，使用多目標演算法生成平衡方案，使能耗成為編譯決策參數。
- 3) **自我調整優化層**：結合歷史資料和任務特徵，系統識別能效模式，在後續執行中優先選擇最優配置。編譯器因此具備學習與記憶能力，形成持續優化機制。

這種結構讓編譯系統從靜態優化器變為能效智慧體。它成為軟硬體協調的主動核心。對於昇騰 CANN，這意味著編譯系統不再是工具鏈元件，而是晶片能效控制中心。通過與功率管理單元 (PMU) 和監控模組連接，編譯器可在任務級別調整運算元並行度和功率分配，實現“軟體控制硬體”的動態優化。

7 結論

本研究以昇騰 CANN 為研究物件，對圖優化編譯的理論邏輯和能效機制進行了系統分析。通過對計算圖結構、運算元融合、記憶體調度和硬體協同的分析，揭示了 CANN 以圖優化為核心驅動力的設計思路。其優化邏輯體現出從演算法語義到硬體執行的映射關係，通過靜態圖重構和資源約束分析，實現軟硬體間的高效配合。這種機制的重點不在於單一演算法性能，而在於利用編譯系統把能效約束納入 AI 計算的結構框架。

在理論部分，本研究提出了計算圖優化與能效建模的分析框架。該框架基於圖論和多目標優化理論，將執行代價、頻寬使用和能耗指標放入同一模型，用來解釋 CANN 在運算元級與系統級的能效調度邏輯。這種框架不同于傳統的性能優先編譯方式，更關注約束下的平衡，使能效優化變成可建模、可分析的過程。這一方法為國產 AI 晶片編譯系統的優化研究提供了理論支援。

在系統比較中，通過分析 TensorRT、TVM 和 NeuWare 等框架，可以看到國產編譯系統的主要取向。以硬體特徵為核心的定制化優化在性能上有明顯優勢，但通用性和生態相容性較

弱。這種現象不是技術缺陷，而是產業階段的特徵。未來的方向是尋求性能與開放性的平衡，使編譯系統既能體現硬體特徵，又能支援多晶片協同與生態統一。

研究仍存在局限。昇騰 CANN 的部分機制未公開，特別是調度與功率約束模型的細節，公開資料有限，理論驗證仍有不確定性。能效模型的驗證主要依賴模擬環境，實際資料的系統分析需要進一步補充。

未來的圖優化編譯將進入能效自調節與智慧優化階段。編譯器不再只是性能工具，而是能效智慧體，負責算力調度、能量分配和模型自我調整。對中國而言，國產 AI 編譯系統的完善不僅是技術問題，也是算力自主的關鍵路徑。是否能在圖優化層面建立制度化、開放化和智慧化體系，將直接影響 AI 計算生態的競爭力。

參考文獻：

- [1] Chen T, Moreau T, Jiang Z, Zheng L, Yan E. TVM: An Automated End-to-End Optimizing Compiler for Deep Learning[C]// Proceedings of the 13th USENIX Symposium on Operating Systems Design and Implementation (OSDI '18). USENIX Association, 2018: 578-594.
- [2] 張芳芳. 基於 CNN 加速器的深度學習編譯器設計與實現[D]. 西安: 西安電子科技大學, 2020.
- [3] 《華為研究》編輯部. 科學與工業中的 AI 應用及其前景[J]. 華為研究計算專刊, 2024(6): 12-25. 深圳: 華為技術有限公司.
- [4] Ding Y, Yu C H, Zheng B, Liu Y, Wang Y, Pekhimenko G. Hidet: Task-Mapping Programming Paradigm for Deep Learning Tensor Programs[EB/OL]. arXiv:2210.09603, 2022.
- [5] Furutanpey A, Walser C, Raith P, Frangoudis P A, Dustdar S. Leveraging Neural Graph Compilers in Machine Learning Research for Edge – Cloud Systems[EB/OL]. arXiv:2504.20198, 2025.

版權聲明

© 2025 作者版權所有。本文依據“知識共用署名 4.0 國際授權合約”（CC BY 4.0）以開放獲取方式發佈。該許可允許使用者在任何媒介中自由使用、複製、傳播與改編文章（含商業用途），惟須明確署名原作者及出處，並注明所作修改（如有）。完整協議詳見：<https://creativecommons.org/licenses/by/4.0/deed.zh-hans>

出版聲明

所有出版物中的陳述、觀點及資料僅代表作者及供稿者個人立場，與 Brilliance Publishing Limited 及/或編輯人員無關。Brilliance Publishing Limited 及/或編輯人員對因內容所提及的任何理念、方法、說明或產品所導致的人身或財產損害概不負責。