

澳門城市大學 榮譽班啟動儀式

2024年9月6日



澳門城市大學榮譽班
Classe de Honra · Universidade da Cidade de Macau
Honors Class · City University of Macau

年度報告

2024/2025學年



澳門城市大學
Universidade da Cidade de Macau
City University of Macau

榮譽班（Honors Class）簡介

澳門城市大學於 2024/2025 學年隆重推出學士學位課程榮譽班。第一屆以數據科學學院為試點單位，在計算機科學學士學位課程及智慧科技與服務學士學位課程大一至大三的學生中進行成員選拔。榮譽班打造創新人才培養模式，基於清華大學錢學森班 15 年培養經驗，澳城大聯合深圳零一學院，攜手院士和國內外專家大導師團隊，共同為數據科學等多方向打造創新人才培養新模式。

依榮譽班特點而全新設計的進階式精深學習體系，從初級科研訓練開始，到自主科研訓練和高階海外實習，不斷激發學生內生動力與熱情，培養綜合創新能力。在這個過程中，學生將循序漸進，不斷受到新知識、新挑戰的洗禮，提升創新綜合素養。同時，來自全球的大導師團隊及菁銳導師，會陪伴心懷夢想、勇於創新的學生一起成長。

2024/2025 學年榮譽班學生人數及專業概覽

專業名稱	學生人數		
	大一	大二	大三
計算機科學	4	3	0
智慧科技與服務	11	10	2
總人數：30			

榮譽班學子的五個維度+一個跨文化溝通能力的基本特徵：

- **內生動力**：對科研和創新具有著迷般的興趣，具有主動性、自驅力，投入度高；
- **開放性**：好奇心強，視野開闊，敢於突破常規，喜歡嘗試新

事物、新想法；

- **堅毅**：對感興趣的領域全神貫注、不屈不撓、克服困難、擁抱失敗；
- **智慧**：基礎知識紮實，學習能力、自我調節能力和創造能力強；
- **領導力**：責任擔當、奉獻精神、資源整合和團隊合作能力強；
- **跨文化溝通能力**：良好的英文表達能力，在具有文化自信的同時能審視自己的文化觀點，能站在他者的角度和立場理解對方，瞭解並包容不同文化背景人際相處上的異同。

榮譽班課程體系

榮譽班課程體系簡介

1) 榮譽班--挑戰性問題 (Honor Class X-Idea)
2) 榮譽班--初階研究訓練 (Honor Class pre-SRT)
3) 榮譽班--高階研究訓練 (Honor Class adv-SRT)
4) 榮譽班--自主研究訓練 (Honor Class ORIC)
5) 榮譽班--海外研究訓練 (Honor Class SURF)

榮譽班課程體系進展情況

目前，“榮譽班-挑戰性問題 (Honors Class X-Idea)”的訓練已正式結束，結題報告會分為 2024 秋季學期及 2025 春季學期兩期，皆已完成學生報告。下一學年將逐步銜接“榮譽班-大學生研究訓練 (Honors Class SRT)”系列訓練。“X-Idea”訓練期間，榮譽班配備一系列的學生活動，如學術講座、主題研討會、科創類企業及機構的探訪、產學研項目、暑期挑戰營等，取得了一定的學術成果。

2025/2026 學年榮譽班將繼續深化各類學生活動，與榮譽班 SRT 課程體系相配合、相輔相成。透過更精準的資源投入與活動設計，有效提升學生的科研素養與實踐效能，為其未來的學術發展或職涯競爭力奠定更堅實的基礎。

榮譽班舉辦活動情況

主題講座

鄭泉水院士《愉快的學與教》



鄭建華教授 《數學教育之思考》



王威教授、丁文伯教授 《關於科研的那些事兒》（線上）



清华大学深圳国际研究生院
Tsinghua Shenzhen International Graduate School

双螺旋中心
Center of Double Helix
0-1 Innovation & Innovators

X-share系列活动

关于科研的那些 事儿



王威

哈尔滨工业大学（深圳）教授、博导；
研究方向：微纳米机器人、
智能仿生材料的交叉学科。

分享主题：
微纳米机器人与材料科学



丁文伯

清华大学深圳国际研究生院副教授；
发展规划办主任；
研究方向：机器人触觉感知
及具身智能等。

分享主题：
视觉与触觉的交融
——面向复杂目标的感知与抓取技术

活动时间：

4月30日（周三）

10:00-11:30 王威教授主题分享

（哈工深H栋514室）

13:00-14:00 校园参访、探索

（大学城）

14:00-15:30 丁文伯教授主题分享

（清华深研院信息楼501室）

實地探訪

2024年9月28日，榮譽班第一屆全體30名學生在榮譽班導師的帶領下開展了為期兩日的深圳科創企業、研究院探訪交流活動。交流團走訪了深圳理工大學、中國科學院深圳先進技術研究院、深圳零一學院和華大基因，激發了學生的創新意識和解決問題的科研能力。



學術研討會

《數學教育之思考》主題研討會



《Revealing Causal Information from Data》主題研討會



學生座談會

本學年共舉辦兩個場次的學生座談會，總參與人次為 60 人次。2024 秋季學期場次座談會主要聚焦澳城大榮譽班現狀；聽取榮譽班學生對於榮譽班期待、平衡學業與科研的疑惑以及長週期培養意願的反饋。結合同期舉辦的榮譽班導師座談會，榮譽班工作組與零一學院指導專家組作出了下一步的學生培養行動計劃。

2025 春季學期場次以小組形式舉行，每組安排 3-5 位同學參與，還邀請了教育學院黃碧雲老師的教育專業觀察組共同參與並給出意見。座談會結束後，榮譽班工作組依據每小組得到的反饋撰寫了報告書，旨在優化榮譽班培養模式、為榮譽班的學生成長提供參考依據，以助力榮譽班的科研產出可以進一步的提升。

2024 秋季學期場次



2025 春季學期場次



學期末匯報會

本學年共舉辦 2024 秋季學期、2025 春季學期兩個場次的期末匯報會，總參與人次為 60 人次。同學們依次上臺展示本學期的研究成果，主要圍繞“本學期的學習及科研進展”、“下學年整體的學習及未來科研計劃”、“當前面臨的挑戰及所需的支持”三個方向展開。在 2025 春季學期場次中，還邀請了零一學院 X-perience 項目主任李莉老師介紹零一學院長週期培養計劃的核心理念及校企合作生態的打造。零一學院依託深圳的產業優勢，已與比亞迪、華為、韶音科技等知名企業建立深度合作，形成產學研緊密結合的創新培養體系。值得關注的是，這一資源網絡已向澳門城市大學榮譽班開放，為榮譽班學生提供更多企業參訪、聯合課題研究等實踐機會，助力學生拓展產業視野並探索個人科研方向。

榮譽班產學研項目

本學年榮譽班聯合澳門產學研合作促進會，在孫沛執行副院長的帶領下走進本地校園，在同善堂中學開展了兩期科創培訓課程，每期課程為期四周。課程內容包括人工智能概覽、科學實驗設計基礎，以及文獻資料查閱與精讀等。

此次產學研項目中，榮譽班共有 12 人次同學參與並擔任課程主講人，為同善堂學生帶來人工智能領域的研究與創新經驗。此舉不僅豐富了學員們的知識儲備，還為他們未來的學習與創新奠定了堅實基礎。也使所有參與者——無論是負責授課的榮譽班同學還是同善堂中學的學生，都在教授課程或學習過程中受益匪淺，對人工智能的可能性有了更深刻的理解與思考。



暑期學校資源

“X-Challenge” 暑期創新挑戰營

零一學院主辦的“X-Challenge”暑期創新挑戰營是集創新人才與科研訓練於一體的特色人才培養項目,面向榮譽班大一、大二學生進行選拔。挑戰營採取“以學術為中心的研究性學習”教育模式,讓學生可以組建跨學科的研究小組挑戰最前沿的問題並提出原創想法,使得“傑出創新導師、天賦異稟的學生、重大原創問題”得以交匯、碰撞和激發。

此次挑戰營也是澳城大榮譽班繼“X-Idea”課程(挑戰性問題課程)結課後,對榮譽班學子的一次實踐性挑戰。進入挑戰營的榮譽班同學將完整經歷“發現問題-分析問題-解決問題”的實踐性質的科研閉環,將“X-Idea”課程中學習到的各種理論與科研技巧靈活運用,實現科研能力的一次躍遷,為未來的學術深造與職業發展奠定堅實基礎。

此次挑戰營榮譽班共4位同學接收到挑戰營的錄取通知,截止目前已有兩位榮譽班同學進入第一期挑戰營進行科研挑戰訓練。



“X-Fusion” 全球創新者聚變大會人才選拔活動

“X-Fusion”是零一學院主辦的另一項暑期科研資源，面向榮譽班高準大三、準大四學生進行選拔。依託“X-Challenge”中通過精心打磨提煉而發佈的重大挑戰問題，結合”X-Idea”中大導師與學生的直觀深入的瞭解和交流，完成學生與大導師的雙向匹配。

榮譽班學生有機會與來自全球的頂尖科學家團隊、頭部可以寄企業面對面、高密度的探討一批前沿重大挑戰問題，找到未來研究及深造方向。同時，學生可以與 100+來自全球頂尖高校的志同道合的創新學子頭腦風暴、組隊研究、尋找重大挑戰問題的顛覆性解決方法。

目前，該活動仍在報名過程中，未能統計實際的入選榮譽班學子人數。



榮譽班導師培養

本學年導師培養緊扣創新人才培育核心，通過三階段系統化培養全面提升導師團隊素質：攜手零一學院專家深化科研指導實務交流，引入多維創新評價體系重構人才培育路徑，並推動兩屆榮譽班導師聯合培訓強化協同效能與前沿視野，為塑造榮譽班鮮明特色人才培養模式奠定堅實基礎。

2024 榮譽班導師與零一學院專家導師見面會

2024 年秋季學期期間，零一學院總班主任、E-SRT（Enhanced Student Research Training）課程負責人任建勛，零一學院創新系統工程部負責人饒軍民兩位專家導師來校進行指導。榮譽班導師代表敬馮時老師、田洋老師與零一學院的導師進行了面對面的座談，從如何培養學生創新意識、鼓勵學生創新以及投入科研、現在榮譽班面臨的一些疑惑和困境等各方面出發，專家與榮譽班導師利用這個機會進行了深入交流。任建勛老師強調榮譽班導師應注重激發學生的科研興趣與探索精神，分享了有效指導學生開展科研訓練的實務經驗，為導師們未來的教學設計提供有益參考。

2025 春季學期末導師培訓

為進一步總結澳門城市大學榮譽班在 2024/2025 學年度的教學經驗與成果，城大榮譽班於學期期末在氹仔校區舉辦導師培訓及座談會，零一學院總班主任、E-SRT（Enhanced Student Research Training）課程負責人任建勛，零一學院創新系統工程部負責人饒軍民，零一學院科學創新部零一班合作項目負責人蔡姿雲，零一學院 X-perience 學生管理主任李莉等專家參與指導，共同探討創新人才培養模式並檢視學生科研進展。

在導師培訓及座談會上，深圳零一學院創新教育系統工程研究中心負責人饒軍民以“大中貫通多維創新人才評價模式及實踐”為主題，系統介紹了五維評價模型及中學-大學-產業一體化貫通式評價體系，該模式突破傳統的單一學術評價框架，從創新思維、實踐能力、跨學科素養等多維度全面評估學生潛力，為榮譽班導師提供了科學化、系統化的人才培養思路。在隨後的座談環節，與會者圍繞課程體系優化、學生科研熱情激發、競賽指導策略及新一屆榮譽班學生選拔機制等議題展開熱烈討論。

未來，榮譽班將進一步落實榮譽班的科研培訓計劃，推動榮譽班形成更鮮明的特色人才培養模式。

2025 暑假第一屆、第二屆導師聯合培訓

為強化導師梯隊協同發展，2025 年暑期啟動首屆跨屆導師聯合培訓“拔尖創新人才聯合培訓”，培訓時間為 7 月 26 日至 29 日。此次培訓共推送 6 位導師參與培訓，其中第一屆導師 3 名、第二屆導師 3 名。屆時將由孫沛執行副院長帶隊前往深圳參與為期 4 日的培訓。

本次導師聯合培訓獲得大學全力支持，導師們將在零一學院南山總部與來自各大知名院校的導師一同參與主題式討論、創新人才培養體系整體架構與邏輯理念主題學習、人才培養方案落地工作坊、與院士及知名教授一起參與信息與智能學域以及生命與健康學域 X-C 大問題發佈等，全面拓展導師視野。

榮譽班學術成果

2024/2025 學年度學術成果包括學術論文的發表及學生的研究方向、學生學年獲得的各類榮譽、參加科研相關的學術會議及參加的科創競賽活動等。

榮譽班現有研究方向

一、 大語言模型與生成式人工智慧方向

- 生成式大模型
- 大語言模型性能研究
- 大模型和自然語言處理
- 機器遺忘大模型 vlm
- AI 幻覺

二、 智慧醫療與醫學人工智慧方向

- 智慧醫療
- 多模態 VQA（醫學領域）
- 醫學心理學電腦視覺深度學習

三、 人工智慧教育應用方向

- AI 與教育
- 推薦演算法-基於知識圖譜的教育推薦
- AI 學生心理健康

四、 電腦視覺與深度學習技術方向

- 電腦視覺，複雜網路影響力最大化
- 神經網路視覺演算法 YOLO
- 深度學習，大資料處理
- 深度學習的安全方向 剪枝方面

五、 特色前沿與交叉應用方向

- 寬度強化學習在無人艇中的應用
- AI 情感陪伴

榮譽班學生發表的學術論文

2024/2025 學年榮譽班共有 6 位同學撰寫論文並成功發表。這些成果不僅體現學生科研能力經過一年的科研訓練後的逐步提升，更彰顯榮譽班「以學術驅動創新」的培養理念，為拔尖人才梯隊建設提供實證支撐。具體論文信息如下：

論文名稱	撰寫者姓名（此處只列明榮譽班學生）	指導老師
Learning to Program with VEX Robotics: Enhancing Computational Thinking and Problem-Solving Skills in K-12 Education	麥標昕	黃碧雲

Quantifying AI Hallucination Severity: A Multicontext Evaluation Framework for LLM Reliability	何瀚	張樂峰
Can LLMs Assist Computer Education? An Empirical Case Study of Deepseek	高晨、肖東福	劉馳
From Thinking to Output: Chain-of- Thought and Text Generation Characteristics in Reasoning Language Models	許振豪、陳溢釗	朱天清

榮譽班學生學年內獲得的榮譽獎項

榮譽名稱	頒發單位
澳門高等院校優秀學生獎學金	澳門基金會
攝影作品人氣獎	澳門城市大學
圖書館攝影比賽三等獎	澳門城市大學

榮譽班學生學年內獲得的競賽類獎項

競賽名稱	獎項	頒發單位
華為 ICT 大賽 2024-2025 澳門站	三等獎	華為公司
“創新、創意及創業”挑戰賽（校級賽）	一等獎	全國電子商務產教融合 創新聯盟與西安交通大學
第十五屆 MathorCup 數學應用挑戰賽	三等獎	中國優選法統籌與經濟 數學研究會
正大杯第十五屆全國大學生市場調查大賽	二等獎	中國商業統計學會

正大杯第十五屆全國大學生市場調查大賽	一等獎	中國商業統計學會
“城市杯”英語演講比賽	季軍	澳門城市大學

榮譽班學子本學年感想精選

“非常榮幸能成為澳門城市大學首屆榮譽班的一員。在過去的一年裡，這個平臺為我提供了寶貴的科研資源與支援，使我得以潛心學習、參與競賽並嘗試科研實踐，這正是我加入前所熱切期盼的機會。

榮譽班為我們配備了實力雄厚、認真負責且充滿熱情的導師團隊和班主任。在他們的悉心指導和鼎力支持下，我們得以在科研道路上勇往直前。作為第一屆學員，最初我確實對這個班級能否成功運行心存疑慮。然而，隨著時間推移，我深切感受到老師們為榮譽班的建設所付出的巨大努力，他們積極摸索、不斷探索，致力於完善各項制度。此外還密切關注每位同學的動態，並定期組織研討會，讓大家交流心得、分享經驗，從而持續優化榮譽班的建設。在高效的安排下，我們很快完成了導師分配。更令我驚喜的是，短短幾個月，在導師劉馳的指導下，我作為第一作者成功發表了第一篇會議論文。能在大學二年級就取得這樣的成果，其速度之快著實令我感到意外與振奮。這無疑是榮譽班給予我的一個里程碑式的收穫。

我深信，在學校的持續投入和老師們的共同努力下，榮譽班必將日益完善，彙集更多懷揣夢想的學弟學妹。作為其中的一員，我倍感自豪，並將以此作為動力，在未來的學術道路上勇攀高峰。”

“本學年，除扎實掌握課內知識外，我還系統學習了機器學習基礎課程，不僅成功複現了相關專案，還學習了學術報告的撰寫方法。期間，我參加了“正大杯”第十五屆全國大學生市場調查與分析大賽，積極與經驗豐富的參賽同學交流學習。在課題組中，我們通過定期組會督促學習進度、及時解決疑難問題，為我的學習指明了方向。此外，我與課題組同學一同刷題，並定期開展線上習題講解會對自己寫過的習題進行講解，提升了我的程式設計能力。而暑期實習則讓我將理論知識應用於實踐，進一步增強了動手實操能力。”

“這學期我跟著老師開始進行實操方面的學習，修改調試別人的代碼，並要在功能效果上面做出一些創新，我覺得這個過程非常的有意思。我學習到了很多知識，有技術方面的也有思維方面的，也激發了我對我的專業的興趣，讓我全方面有了很大的提升。這學年要同時兼顧課內和課外的作業，也讓我對時間分配管理更加上心，也認識到了時間有效利用的重要性！非常開心我能夠加入榮譽班！”

“本學年的科研學習主要在基礎方面得到了訓練，如電腦視覺的基礎課程，論文查找、閱讀、梳理等，並且開始進行文獻綜述的寫作工作。這部分的學習也對我本科的課程內容學習頗有助力，學會主動深入探索和研究問題。

但與此同時，本學年由於課程和個人考試進度，對於論文和問題研究的工作時斷時續，我也在這個過程中學會了積極調整個人計畫和學習狀態。非常感謝導師和榮譽班各位老師的教導和引領，希望在下一學年能夠在所研究的問題上得到大幅進展，並且發表論文。”

“在本學期中課內學業較為繁忙，加之需要準備雅思考試、全澳

大學生錦標賽，所以沒有在科研方面取得太大的進展，只是跑了一些模型、閱讀和複現了一些論文。在下學期的學習中，課程量會減少，我可以有更多的時間和精力放在科研上，希望在下學期的科研學習中能夠有所產出。

除此之外我認為我在實習時期收穫較多。在樂動機器人公司實習期內，我學到了很多關於電腦視覺方向的内容（資料清洗、影像處理、yolo 演算法等），並與同事一起訓練了一些圖像識別的腳本（主要是二分類，提升其識別的精確度）來提升工作效率。我認為在這期間所積累的學習和工作經驗將會對我未來的學習有所幫助。”

附錄：學生論文及首頁

1. Learning to Program with VEX Robotics: Enhancing Computational Thinking and Problem-Solving Skills in K-12 Education

GCCCE2025

運用 VEX 機器人學習編程：提升 K-12 教育中的計算思維與解決問題能力 Learning to Program with VEX Robotics: Enhancing Computational Thinking and Problem-Solving Skills in K-12 Education

Iek Chong Choi¹, Biyun Huang^{1*}, Qiqi Chen¹, Biaoxin Mai²

¹School of Education, City University of Macau, Macao SAR, China

²Honors Class of Faculty of Data Science, City University of Macau, Macao SAR, China

*byhuang@cityu.edu.mo

【摘要】 隨著科技的迅速發展所帶來的技術創新與數位化轉型的需求不斷提升，K-12 階段的資訊科技教育變得愈加重要。然而，對初學者而言，學習程式設計具有挑戰性，因為它需要邏輯應用與問題解決等複雜的認知能力。許多學生即使在入門課程中也感到困難。本研究訪問了兩位中學階段的電腦和機器人教師，以瞭解了他們對於將 VEX 機器人運用于編程教學中的實踐經驗與看法。訪談結果表明，VEX 機器人，為提升學生的計算思維及解決問題能力提供了有效途徑。而且，通過強調創新與互動的教學方式，有助於提升學生的學習動機與參與度。

【關鍵字】 K-12 教育；學習動機；計算思維；認知負荷；VEX 機器人

Abstract: The importance of computer science education in K-12 has increased dramatically in response to the demand for technological innovation and digital transformation brought about by rapidly advancing computational technology. However, mastering programming can be challenging for beginners, requiring complex cognitive abilities such as logical application and problem-solving. Many students struggle, even in introductory courses. This study interviewed two K-12 STEM teachers to explore their practical experiences and perspectives on integrating VEX robots into programming instruction. The interview results indicate that VEX robots effectively enhance students' computational thinking and problem-solving abilities. Furthermore, the strategy promotes innovation and interaction, which increases students' learning motivation and engagement.

Keywords: K-12 Education, Learning Motivation, Computational Thinking, Cognitive Load, VEX Robotics

1. 前言

隨著未來科技發展驅動社會對技術創新與數位轉型需求的增加，K-12 年級階段的資訊科技教育重要性亦日益提升 (Chen et al., 2017)。21 世紀的來臨預示著計算思維 (Computational Thinking, CT) 已成為一項關鍵能力，不僅在計算機科學領域中舉足輕重，在其他領域亦具深遠影響。CT 強調系統設計、問題解決能力與邏輯推理等核心技能的培養 (Wing, 2006)，構成了演算法與程式邏輯的基礎，對於開發高效能電腦應用程式至關重要 (Aho, 2012)。

然而，培養學生的計算思維能力是一項充滿挑戰的任務。研究指出，儘管學生能依照教師的示範撰寫程式，卻常無法理解程式結構背後的原理 (Lye & Koh, 2014)。當被要求獨立開發程式時，學生往往難以正確完成，亦無法有效運用計算思維來解決問題。Guzdial (2015) 指出，學生在程式設計中所面臨的困難涵蓋錯誤觀念、練習不足及學科本身的複雜性。

初學者常覺得程式設計困難且複雜 (Sáez-López et al., 2016)。許多學生即使在入門課程中也面臨困難，因為掌握程式語言對他們而言具有挑戰性 (Mow, 2008)。同時，對於文字程式語言的文法理解及指令輸入，學生普遍感到困難，這些問題難以完全消除 (Wae-Shik, 2006)。傳統教學方法已無法有效應對多元學生的學習需求 (Papadakis & Kalogiannakis,

2. Quantifying AI Hallucination Severity: A Multicontext Evaluation Framework for LLM Reliability

Quantifying AI Hallucination Severity: A Multicontext Evaluation Framework for LLM Reliability

Han He¹ 

City University of Macau
Avenida Padre Tomás Pereira Taipa, Macau 999078, China
D23090103904@cityu.edu.mo, hhisurdad@163.com
<https://cityu.edu.mo/zh/>

Abstract. When used, large language models (LLMs) frequently generate hallucination statements that lack factual accuracy, are logically incoherent, or are contextually misleading. Such hallucinations can occur when a model produces content that is not part of the user input but has been hallucinated in contrast to the produced context that also contradicts the world knowledge we know to be true [1]. Though it is a topic of increasing concern, literature to date has been dominated by work on large-scale proprietary models; the space of small, locally deployable open-source models has received scant attention.

In this paper, we propose developing a systematic approach to evaluating locally deployable LLMs, given that these locally deployable LLMs are susceptible to data homogenization and other factors. We present a three-category evaluation regime: (i) fact-based question answering; (ii) multi-step reasoning tasks, and (iii) adversarially induced queries. Furthermore, we propose three new metrics for quantifying hallucination severity: the factual error rate, the logical contradiction rate, and the accuracy score. This framework can be used as a standardized benchmark for the reliability of LLMs, and it enables quantitative comparison of models' hallucination tendencies by different model scales.

Keywords: AI Hallucination Quantification · Multicontext LLM Evaluation · Factual Consistency Metrics.

1 Introduction

The integration of large language models (LLMs) into high-stakes domains has raised significant concerns regarding their reliability. Hallucinations—factually incorrect or logically inconsistent outputs—pose particularly severe risks in mission-critical fields such as medical, banking, finance, legal, and clinical environments, in which error or incorrect information may lead to profound consequences [2]. Of course, these issues impact all LLMs, but medium-sized open models (6B–70B

3. Can LLMs Assist Computer Education? An Empirical Case Study of Deepseek

Can LLMs Assist Computer Education? an Empirical Case Study of DeepSeek

Dongfu Xiao¹*, Chen Gao¹*, Zhengquan Luo¹, Chi Liu¹✉, Sheng Shen²

¹ Faculty of Data Science, City University of Macau, Macao SAR, China

² Design and Creative Technology Vertical, Torrens University Australia, NSW, Australia

* Equal contribution

✉ Corresponding author: chiliu@cityu.edu.mo

Abstract. This study presents an empirical case study to assess the efficacy and reliability of DeepSeek-V3, an emerging large language model, within the context of computer education. The evaluation employs both CCNA simulation questions and real-world inquiries concerning computer network security posed by Chinese network engineers. To ensure a thorough evaluation, diverse dimensions are considered, encompassing role dependency, cross-linguistic proficiency, and answer reproducibility, accompanied by statistical analysis. The findings demonstrate that the model performs consistently, regardless of whether prompts include a role definition or not. In addition, its adaptability across languages is confirmed by maintaining stable accuracy in both original and translated datasets. A distinct contrast emerges between its performance on lower-order factual recall tasks and higher-order reasoning exercises, which underscores its strengths in retrieving information and its limitations in complex analytical tasks. Although DeepSeek-V3 offers considerable practical value for network security education, challenges remain in its capability to process multimodal data and address highly intricate topics. These results provide valuable insights for future refinement of large language models in specialized professional environments.

Keywords: Large Language Models · Computer Education · Empirical Evaluation

1 Introduction

Recent advancements in large language models (LLMs) have opened substantial opportunities to transform professional education in computer networking and security. These models exhibit advanced capabilities in processing and articulating intricate technical concepts[20,21], making them potent educational tools that can complement conventional learning methods[18]. Their capacity to produce precise, context-aware explanations of networking and security principles[11], coupled with the generation of tailored instructional content[14], establishes LLMs as valuable resources for developing professional expertise. Although initial

4. From Thinking to Output: Chain-of-Thought and Text Generation

Characteristics in Reasoning Language Models

From Thinking to Output: Chain-of-Thought and Text Generation Characteristics in Reasoning Language Models

Junhao Liu¹[0009-0007-9353-6929], Zhenhao Xu¹[0009-0002-1834-5181], Yuxin Fang¹[0009-0006-5220-3274], Yichuan Chen¹[0009-0006-4992-5559], and Wenhan Chang^{*2}[0000-0003-3350-5171]

¹ Faculty of Data Science, City University of Macau, Macau SAR, China
611yuyi@gmail.com, {D23090100776,D24090103371,D24090150413}@cityu.edu.mo

² School of Information Engineering, Zhongnan University of Economics and Law,
Wuhan, China
changwh530@gmail.com

Abstract. Recently, there have been notable advancements in large language models (LLMs), demonstrating their growing abilities in complex reasoning. However, existing research largely overlooks a thorough and systematic comparison of these models' reasoning processes and outputs, particularly regarding their self-reflection pattern (also termed "Aha moment" [3]) and the interconnections across diverse domains. This paper proposes a novel framework for analyzing the reasoning characteristics of four cutting-edge large reasoning models (GPT-o1 [10], DeepSeek-R1 [3], Kimi-k1.5 [13], and Grok-3) using keywords statistic and LLM-as-a-judge paradigm. Our approach connects their internal thinking processes with their final outputs. A diverse dataset consists of real-world scenario-based questions covering logical deduction, causal inference, and multi-step problem-solving. Additionally, a set of metrics is put forward to assess both the coherence of reasoning and the accuracy of the outputs. The research results uncover various patterns of how these models balance exploration and exploitation, deal with problems, and reach conclusions during the reasoning process. Through quantitative and qualitative comparisons, disparities among these models are identified in aspects such as the depth of reasoning, the reliance on intermediate steps, and the degree of similarity between their thinking processes and output patterns and those of GPT-o1. This work offers valuable insights into the trade-off between computational efficiency and reasoning robustness and provides practical recommendations for enhancing model design and evaluation in practical applications. We publicly release our project at: <https://github.com/ChangWenhan/FromThinking2Output>

Keywords: Reasoning Language Models, Reasoning Process, Reasoning Pattern Analysis

* Corresponding Author

聯絡我們

辦公地點：澳門氹仔徐日昇寅
公馬路大豐樓二樓T221室

辦公時間：周一至五 09:00 -
12:45、13:45 - 18:00

電郵：honorsclass@cityu.edu.mo



澳門城市大學榮譽班
Classe de Honra - Universidade da Cidade de Macau
Honors Class - City University of Macau

