粵語語料庫建設與大模型評測重點實驗室成立

2024 年 11 月 14 日，大模型與語言資源學術研討會暨粵語語料庫建設與大模型評測重點實驗室啓動會在廣州大學（黄埔校區）舉行。此次大會由廣州大學與廣州市社會科學界聯合會聯合主辦，彙聚了來自廣東省、市的政府領導、學術委員會專家學者以及産業企業代表等約百餘位嘉賓，旨在推動語言資源的數字化與大模型應用的多語言服務化，標志著廣東省在粵語語料庫建設方面邁出了重要一步。

“人文科學 + 科技創新”傳承保護粵語，探索中國特色、嶺南風格的城市文化傳承、保護與發展之路

粵語是嶺南文化最重要的基因和最具特色的要素，爲全面貫徹落實黨的二十届三中全會精神，弘揚廣州歷史文化，傳承廣州千年文脈，廣州大學與廣州市社科聯經過多次調研、多方論證，决定以人工智能時代粵語語言的傳承發展爲著力點，合作共建“粵語語料庫建設與大模型評測重點實驗室”，期望通過多學科交叉融合創新研究，探索具有中國特色、嶺南風格的城市文化傳承、保護與發展之路。

“粵語作爲中華文化的瑰寶，承載著粵港澳大灣區深厚的歷史文化底蘊。我們有責任，也有義務，利用現代科技手段，保護和傳承好這一寶貴的文化遺産。”廣州大學張其學副校長在致辭中指出，“粵語語料庫建設與大模型評測重點實驗室”（以下簡稱“實驗室”）的成立，是廣州大學服務國家戰略、響應時代號召的具體行動，是廣州大學在跨學科建設中邁出的重要一步。實驗室不僅能推動粵語文化的數字化傳承，更能在粵港澳大灣區乃至全球的文化交流與合作中發揮積極作用，還將成爲廣州大學跨學科發展中的重要學術平臺。期待實驗室在未來能够結出豐碩的成果，爲中華文化的傳承與發展，爲粵港澳大灣區的繁榮與進步，爲推動國家科技創新、提升文化自信，貢獻智慧和力量。

實施哲學社會科學創新工程，全面推進學科體系、學術體系和話語體系建設和創新，積極探索人文科學與科技創新相融合的路徑，奮力在構建中國哲學社會科學自主知識體系上展現新擔當新作爲。”廣州市社科聯主席崔頌東指出，此次與廣州大學合作共建粵語語料庫建設與大模型評測重點實驗室，是打造跨學科協同創新科研平臺、緊抓哲學社會科學創新“新機遇”、錨定學科交叉融合破題大模型“新賽道”的重要舉措。他强調，要扎扎實實把實驗室各項建設落到實處，爲廣州加快實現老城市新活力、“四個出新出彩”，在進一步全面深化改革、推進中國式現代化建設中當好排頭兵，繼續在高質量發展方面發揮領頭羊和火車頭作用貢獻社科力量。

“粵語語料 + 大模型”創新融合，重點實驗室“雙首席科學家”領航探索粵語發展新高地

重點實驗室主任、廣州大學齊佳音教授指出：“當語言遇到大模型，每個國家，每個語種，都要重新審視自己的文化語言戰略。”重點實驗室將通過建立嶺南文化核心語料及細分領域知識庫、粵語數據資産庫、粵語高質量數據集，面向人工智能應用的粵語文化數據資源基礎，形成粵語服務能力以及內容安全合規評測能力，粵語的數據化、資源化、服務化和産業化。

實驗室設立“雙首席科學家”：由中國工程院院士、廣州大學網絡空間安全學院名譽院長、中國電子信息産業集團有限公司資深首席科學家方濱興院士，香港科技大學（廣州）黨委書記、廣州大學二級教授、國家語委科研機構國家語言服務與粵港澳大灣區語言研究中心主任屈哨兵教授分別擔任重點實驗室技術首席專家與語言首席專家，承擔學術指導和諮詢工作，保障重點實驗室學術發展方向。廣州大學張其學副校長和廣州市社科聯崔頌東主席擔任實驗室指導委員會主任，

協調整合各界資源，爲重點實驗室運行提供指導。

“1+1+N”運行機製創新，打造人工智能友好的通用粵語語料庫

重點實驗室的核心任務是建設面向全球、體現粵語文化特色、大規模、多模態、高質量、安全可信、人工智能友好的新型通用粵語語料庫。依托這個新型粵語語料庫可爲粵語語言研究、語言教學、語言智能應用等多種用途領域提供堅實的數據支撐。重點實驗室研究方向包括：粵語文化精神標識體系研究；通用粵語語料庫建設；大模型粵語交互質量評測。

重點實驗室定位爲跨學科的前沿交叉樞紐型研究機構，其組織架構爲“1+1+N”，兩個“1”分別爲市社科聯和廣州大學，“N”爲根據應用研發需求以項目形式組建的若干科研團隊，結合不同研究階段的重點課題和應用場景，吸納高校、科研機構、社會組織和相關企業的專家組成研究團隊，設立開放課題、語料庫聯盟、基地、工作室等，深度參與破解粵語語料庫建設中政産學研用的資源集成難題。

目前，暨南大學漢語方言研究中心、香港科技大學（廣州）方言大模型團隊、華南師範大學文學院、廣州市大灣區現代産業發展研究院等機構已經成爲“1+1+N”機製中的重要機構。2025 年，重點實驗室將在基礎理論、通用語料庫建設、場景語料庫建設等方面開展 13 項研究任務。