

《AI 大講堂》

第八集 - AI 未來十年：香港的機遇與挑戰

視像文字稿

歡迎收看《AI 大講堂》
我是方健儀
曾幾何時
我們和 AI（人工智能）溝通
主要是靠輸入文字或打字
才可以與它聊天
但到今時今日
我們不只用文字
連相片、圖片
AI 都可以理解得到
還會給我們答案
說得好像很簡單
但其實人類和 AI 之間的溝通
每天都有不同的變化
而變得這麼快的話
香港的 AI 技能又可不可以追上
而我們有沒有一些獨特的優勢
在國際 AI 舞台上分一杯羹
這個節目《AI 大講堂》
今天請來的嘉賓
香港理工大學
計算機及數學科學學院副院長
本身也是生成式人工智能
講座教授楊紅霞教授
楊教授多年來從事 AI 的研發
也是致力降低 AI 的研發門檻
很希望 AI 不單止
落入大企業的手中
就連小機構都可以參與得到
楊教授你好
你好
我們知道 AI 無論是技術或應用
都非常廣泛
我們現在說的是「多模態 AI」

即是以前是用文字才能溝通
現在連聲音、相片、影片
它都可以理解
可否跟我們講述「多模態 AI」
對於香港日常市民生活
有何影響及變化
謝謝主持人的問題
首先確實我們過去幾年從大模型
尤其是 ChatGPT 出現之後
我們看到更多的是
「大語言模型」的各種突破
其實「多模態大模型」
在這幾年的發展
相比於「大語言模型」
還是有一定的滯後性的
但是我們想一想
我們周圍所有的各種問題
我們所有的日常生活環境
本質都是一個多模態（大模型）
所以我們要知道一點是
大模型的終極狀態一定是多模態
但是為甚麼「大語言模型」
其實相比於「多模態大模型」
它的進展速度會更快
因為人類經過了大概幾千年
甚至上萬年的演化
語言這件事情
雖然可能你說粵語（廣東話）
我講漢語（普通話）
但其實我們叫 **information**
granularity（資訊粒度）
就是訊息的濃度本身是完全對齊
其實訓練大模型的時候
數據你可以理解它是非常的乾淨
但是訓練「多模態大模型」
主要的數據還是互聯網上叫
Image-Text（圖文）的環繞文本
就是圖文的數據

但這個數據的乾淨程度
其實是遠弱於「大語言模型」
所以這也導致「多模態大模型」
其實發展相比於「大語言模型」
是有一定的滯後性的
但是未來隨著「大語言模型」的
不斷的進展
隨著「多模態大模型」的數據
不斷清洗
以及未來我們的終極目標
一定是個「多模態大模型」
所以接下來「多模態大模型」
一定會不斷的蓬勃
更加蓬勃的發展
大家對多模態的大模型或者 AI
都有一個簡單的理解
可否舉一些很貼地的例子
讓我們了解到多模態的 AI
比純文字的 AI 更加優勝
「多模態大模型」的優勢本身
就是它的覆蓋面會更加的廣泛
可能會應用於日常生活各方各面
就舉個簡單例子
可能你在訂餐的時候
你看的這個帳單
或者是這個菜單本身的語言
我是不認識的
但是可能現在通過拍攝一張照片
手機拍攝一張照片
上傳到各種 Chatbot
（聊天機器人）的軟件
然後給你進行一個自動的翻譯
但是未來我相信
會有一個更加好的入口
這個入口本身其實現在
所有全球的人都是在探索當中的
大家現在可能認為是眼鏡
可能是未來的入口

未來就是你戴一個眼鏡
你看到任何東西
它會自動的給你進行翻譯
然後你的眼球就會進行接收
簡單來講可能你到任何一個地方
雖然我沒有來過這裡
但是馬上會有一個這地方的介紹
各種的翻譯
可能還有我在聽語言的時候
也許我不是聽得很懂粵語
但是這個眼鏡本身可能會
自動的、實時的
給我進行各種的翻譯
其實剛才楊教授你說去點餐
我不懂得看菜牌的外文
其實我經常都會出現這個狀況
所以我都拍照給 AI 之餘
亦都會請它推介一些菜式
有時我們吃完那餐飯
拿了張帳單
亦都拍照給 AI 說
「麻煩你幫我入帳」
處理帳目或者錢銀瓜葛
很多時候，我們市民
都會有這種用到 AI 的狀況
但有沒有一些貼士給大家
用的時候可以更加精準
不會出錯
大家要注意一個點
今天我們這種使用的方式
叫「調用中心化大模型」
API Service 應用程式界面服務
不管是 OpenAI 的 Cloud 雲端服務
還是國內的 DeepSeek 深度求索
像「千問」等等
像「豆包」等等
這叫中心化的調用
API Service 應用程式界面服務

這裡面有一個問題是甚麼呢
我們去問它的所有的數據
其實它們都是會被接收到的
所以這裡有一個
數據隱私（私隱）的問題
所以我覺得一些比較敏感的行業
例如醫療、法律
還有自己個人的一些
非常隱私（私隱）的數據
大家在使用的時候
一定要非常的小心
所以這是第一點
然後第二點就是
目前的這種使用方式
還是一個中心化
叫 **Cloud-based**（雲端化）
例如可能你到了一個
Suburban area（郊區）
就是一個很偏遠的地方
然後 **Wi-Fi** 的上網速度可能很差
那可能在調用的時候
它整個的 **Latency**（延遲）
時長的體感可能會不是很好
其他的方面我是覺得
Again（再次），就是因為
這個大模型能力實在是非常的強
大家還是要更加廣泛的去用
然後探索很多的一些未知
讓我們耳目一新的一些應用
其實今天大模型在訓練的時候
不管是 **Tech Giant**（科技巨頭）
還是 **Startup**
Unicorn（獨角獸初創企業）
他在訓練大模型的時候
使用的數據
主要還是互聯網上
公開可爬取的數據
所以大家會覺得

解決我們日常生活的這些問題
是解決的非常好
但是涉及到一些高精尖
(高級、精密、尖端)的領域
例如像醫療、法律等等
這些數據可能在互聯網上
沒有辦法很好的爬取
其實你在使用它的時候你會發現
它的這個正確率可能就會很差
舉個簡單例子，今天我們至今
沒有看到任何一個大模型可以做
Cancer treatment（癌症治療）
但其實這個是非常有潛力的
不是說模型能力不行
而是這些數據
在互聯網上還沒有辦法獲得
我覺得這是未來一個非常好的
一些落地的可能性
大家現在經常用到不同的 AI
都覺得非常方便
但大家用的時候未必會知道
我們去建構或者訓練 AI 其實很貴
還有背後的成本可能是天文數字
楊教授
可否用一些很簡單的例子去說明
為甚麼開發一個 AI 要這麼貴
其實現在確實是這樣子
因為大家都知道
簡單來說，訓練大模型是非常貴
目前我覺得可能真正的意義上
去訓練一個
Foundation Model（基礎模型）
我們現在所謂的
Centralised（中心化）
Target for
AGI（以通用人工智能為目標）
Artificial General
Intelligence（通用人工智能）

的一個大模型
我覺得幾百億
甚至上千億的投入
我覺得是一件很正常的事情
除了訓練成本之外
其實每天的調用成本
就是大家看到我們自己個人的
是 **individual user**（個人用戶）
大家現在都是免費使用的
但是很多的
business user（企業用戶）
他是要花錢使用的，對吧
但是其實這個叫調用成本
這個調用成本每天的
individual user（個人用戶）
目前的這些大廠
還是這些 **Startup**
Unicorn（獨角獸初創企業）
每天的花費都是巨高(非常高)的
真的有可能是上億的
或者是幾個億的一個
API Service 應用程式介面服務
GPU（圖形處理器）的調用成本
所以整體的成本是非常高的
其實我覺得都不是一個成本問題
從我的角度來講
我覺得未來再過一段時間
去訓練一個非常大的
AGI Foundation
Model（通用人工智能基礎模型）
這件事情，我覺得就不太可能
原因是甚麼呢
核心的主要原因不是算力的問題
也不是人才，也不是技術的問題
我覺得是數據的問題
其實大家在看
我們在訓練大模型的時候
主要還是互聯網公開

可爬取的數據去訓練的大模型
但是大家要知道
包括像谷歌的網頁
可能十年前很多數據已經找不到
它已經下線，被刪除掉
谷歌這個公司它可能會有
但是在互聯網上其實是找不到
這幾年尤其是這兩三年
大家用生成式人工智能越來越多
其實這裡面的很多的數據被污染
怎麼理解這件事情
大家是用了大模型去生成的數據
再去產生很多的 **content**（內容）
那大家想一個問題
人類在演化繁殖的時候
如果是近親繁殖，繁殖的更多
那最後其實生出來的
都是低能兒，對不對
未來大模型如果再這麼演進下去
如果訓練的數據都是互相被污染
互相的近親繁殖
大模型再去訓練一個
Foundation Model（基礎模型）
的能力本身
肯定我覺得也是越來越下降的
所以我個人的判斷
我覺得 **Target for**
AGI（以通用人工智能為目標）的
Foundation（基礎）大模型
中國也好、美國也好
可能就是這幾家了
這是目前的一個階段
但是其實舉個簡單例子
當這個大模型真正的應用在一些
高精尖(高級、精密、尖端)領域
不管是教育也好，還是醫療也好
其實你不需要去使用那麼大的
一個萬億參數

甚至是幾萬億參數
因為最近大家做過一定的測算
就是 GPT-5.5（生成型預訓練
變換模型 5.5）最新的一個版本
可能是 9 萬億參數的一個大模型
這個規模是非常大的
其實你把這個 9 萬億
或者是幾萬億一個大參數模型
在本地化進行一個部署
去解決一個本地化的一個
Domain-specific
problem（特定領域問題）
這件事情其實也是不現實的
所以這也是我們其實
今天在解決的一個問題
第一個問題，本地化的高精尖
(高級、精密、尖端)的數據
不在互聯網上
現在大模型
performance (表現) 是很差的
然後第二，你拿一個巨大的模型
去解決一個很小的問題
其實就是拿了一個大殺器去解決
殺了一個小雞一樣
這個成本，本身也是不合適的
那這個問題我們該怎麼去解決
我覺得我們還是
得到了一些不錯的一些結果
大模型的 Philosophy（哲學）
叫「One for
all（萬能通用一模多用）」
一個大模型甚麼都能解決
如果可能某一個領域的數據過多
可能導致其他領域的能力就會
得到一定的 Trade-off（取捨）
這個問題該怎麼解決
數據該怎麼處理
Data mixing（數據混合）怎做

所以這裡面又花了大量的
GPU（圖形處理器）的算力調整
所以 to summarise (總結)
本身這個訓練
Scaling up（擴大規模）
是一件非常 cost
expensive(成本昂貴)的事
我們休息一下
稍後回來繼續請教楊教授
休息回來
繼續《AI 大講堂》的時間
我們的嘉賓是香港理工大學
計算機及數學科學學院副院長
本身是生成式人工智能講座教授
楊紅霞教授
楊教授你好
如果研發 AI 的成本這麼高
對一些小型研究機構
或中小企業有甚麼影響
最近我們會看到很多中小企業
他們會搭建
就是依據這些不管是開源的
還是閉源的大模型
自己會搭建本地的叫
Agentic framework 代理式架構
簡單來說就是 Agent（代理）的
一個 Workflow（代理工作流程）
然後加速自己的大模型
或者加速整個公司的一個提效
因為我聽說過有一個中型公司
把自己的整個、每一個
Workflow（工作流程）的模塊
給它 Agent 化（代理化）之後
整個公司的效率提升百分之四十
我相信這個事情未來很多公司
都會去採納
所以這是一種
目前大家會廣泛採用的方式

但是 **Again**（再次）這個不解決
說我有領域模型這件事情
因為其實今天大量的中小企業
政府、醫院、機構
它們的數據在互聯網上不能獲取
所以其實這些中小企業也好
這些醫院、這些政府機構也好
他們會直接拿著
這些中心化的大模型進行使用
他們會發現真正的
去解決一個非常複雜的問題
它是解決不了的
但是可以給一些 **regular**(常規)
的 **Workflow**（工作流程）
提速是可以做到的
但是舉個簡單例子
做一個非常複雜的癌症病人診治
它是完全做不到的
所以我覺得這裡面未來
其實是有非常大的空間去做的
這個簡單來說，從我的角度來說
這叫「**AI 原生**」，**AI native**
大家應該知道互聯網的這一代
我們叫 **Cloud native**
就是「雲原生」
「雲原生」是讓互聯網這一代
蓬勃的發展
然後「雲原生」裡面
非常重要的一個基礎設施
叫 **Linux**，「**Linux 操作系統**」
例如我可以在網上採購
然後去打個車（**Call 車**）
直接在手機上進行操作
這個核心的
Contributor（貢獻者）
其實都是「雲原生」
以及後面的 **Linux 操作系統**帶來
我們知道就像你剛才問的問題

接下來這些中小企業一定是希望
最終一定會變成叫「AI 原生」
「AI 原生」不是簡簡單單的說
我今天拿了一個 ChatGPT
然後去寫一個電郵
不是這麼簡單
簡單來說就是今天的「AI 原生」
如果我今天重新再做這個企業
我通過「AI 原生」的思想
這個企業到底該怎麼做
但是我們會發現
這中間現在是缺少一個
infrastructure（基礎建設）
沒有類似像是 Linux 操作系統
的一個存在
所以這也是我們其實覺得接下來
我們自己在非常專注地在做
我們也是覺得這可能是
香港的一個非常好的機會
因為這個操作系統
對於「AI 原生」的操作系統
真正意義上
可以讓各行各業的中小企業
從「AI 原生」
可以 benefit（獲益）
而不單單的就是建一個操作流
這麼簡單的一件事情
可以把這個行業本身徹底的顛覆
我相信這裡是有非常多的機會
另外我知道楊教授
過去這麼多年來都很致力推動
「Co-GenAI」
即是「協作生成式人工智能」
你很想降低研發和訓練門檻
你怎樣令到成本可以大大降低
其實「Co-GenAI」
就是我們可以理解它
就是「AI 原生」時代的

Infrastructure（基礎建設）

操作系統

就類比於說 Linux 作業系統

為 Cloud native「雲原生」

我們今天在做的是

「Co-GenAI Platform」

（協作生成式人工智能平台）

就是「AI 原生」的操作系統

我們做了哪些事情

第一個我們做到了是非常的

Extremely low-resource

（極低資源）去訓練一個模型

因為我剛才已經說了

其實今天大量的中小企業

大量的醫院、大量的政府機構

他們有大量的數據是本地化的

是互聯網上是沒有辦法獲得的

今天大廠它又獲得不了這些數據

所以這裡就是有個 **Gap**（落差）

今天的大模型沒辦法去服務這些

高精尖(高級、精密、尖端)領域

他們自己又想去用這些

真正意義上有自己的領域模型

去打造一個「AI 原生」的企業

他們又做不了這件事情

所以我們今天做的第一件事情

就是 **Extremely**

low-resource（極低資源）

就是你訓練一個模型

可能我只用別人二分之一的資源

甚至是四分之一的資源

裡面我們做到一個叫低比特訓練

叫 **Floating Point**（浮點）

現在的大廠一般用的是

FP16（16 位元半精度浮點數）

或 **BF16**（16 位元大腦浮點數）

這是一個算子精度

我們今天已經完成了端到端的

FP8（8 位元浮點數）
而且我們特別希望
今年我們可以成為全球首發的
FP4（4 位元浮點數）
這個意味著甚麼
極端來講我只用別人
四分之一的算力和存儲
我可以去訓練一個同樣好的
一個領域的模型
所以這是解決的第一件事情
第二件事情，我們為甚麼叫
Collaborative Generative
AI（協作生成式 人工智能）
就是協作式分佈式生成式
人工智能
簡單來說
今天大家在訓練大模型的時候
都是一個大廠做自己的
比如「阿里」（阿里巴巴）
做自己的大模型
不會和任何的人
任何的意義上進行任何的分享
但是就是像我剛才說的
其實今天有大量的領域知識
是不在互聯網上的
你這種中心化的方式
是很難解決這種不同領域的問題
而且也許你很有錢
你可以開一個超大的算力中心
解決這個訓練大模型
但是你永遠不可能把全球最好的
最高精尖(高級、精密、尖端)的
數據放在一個地方
那我們其實今天的解決方案
就像我剛才說的
每一個領域，我以最低的算力
讓你去訓練一個該領域的
一個非常強的領域大模型

那你再把這些領域的模型
Contribute back（回饋）
那可能今天全球
在成千上萬個領域
我把這些模型融合
融合完後，我會得到一個新的
Foundation Model 基礎模型
這個新的 **Foundation**
Model（基礎模型）
會在這成千上萬個領域
比今天 **Centralised**（中心化）
自己做的那個大模型
在這個領域上會做得更好
那意味著我在這成千上萬個領域
我會得到一個真正意義上
全球協作的一個 **Foundation**
Model（基礎模型）
我個人覺得這真正是大家在協作
而不是今天只是 **rich people's**
game (富人的遊戲)
集體的力量是很重要的
現在距離你的目標還有多遠
我們其實現在
已經有了一些不錯的
Proof of Concept（概念驗證）
比如說我們跟中山大學腫瘤醫院
像去年的 **DeepSeek-R1** 也好
還是現在的谷歌的 **Gemini 3**
他們都很快地會調用這種最強的
Centralised（中心化）大模型
放到他們的真實問題當中
但是就像我剛才說的
他們本地化的大量數據
根本不在互聯網上
所以他們的執行成功率
Successful rate（成功率）
只有百分之五十幾
這對醫療來說是完全不可以接受

但是我們其實最近的 13B
就是 130 億的參數模型
遠遠少於 Gemini 3
已經是萬億參數大模型
遠遠少於一個萬億參數大模型
但是在這些高精尖
（高級、精密、尖端）診治領域
我們的準確率是
百分之九十至九十五
現在也是在推進這個臨床實驗
包括也在撰寫
一些 Nature paper
（《自然》雜誌論文）等等
所以我們現在是慢慢的
通過這些非常有挑戰性的
而且不是選那種很簡單的個案
有挑戰性的個案
我們現在也在看教育方面相關的
有挑戰性的個案
去驗證我們整個的
Infrastructure（基礎建設）
是可行的，而且是非常強的
我們在今年大約七、八月的時候
我們也會發佈這樣的一個
Infrastructure（基礎建設）
我們最近也在跟數碼港
也是我們香港的超算中心
可能未來香港會有其他超算中心
我們也會去進行合作
把 Infrastructure（基礎建設）
搭在這些超算中心之上
就是把它們搭上了翅膀一樣
可以更好的給廣大的中小企業
而且我覺得不限於香港中小企業
因為大家也知道
現在國內很多的企業
是通過香港要出海，出海的話
我們怎麼幫他們以最少的成本

能搭上 AI 這輛快速馬車
快速的列車
我相信這個是我們還蠻有信心
可以做成的一件事情
香港有八所大學，也很優秀
但在人才培訓方面
我們有沒有哪一塊
或者哪一方面人才需要加強力度
又或者在香港的中小學
甚至大學方面
我們怎樣配合人才培育
對，我其實最近看到了
香港政府在教育方面
真的是做了非常好的一些措施
比如說最近的中小學教育
STEM（科學、技術、工程、數學）
加強 **STEM** 和 **AI** 學習
然後每所學校給 50 萬港幣
去探索 **AI** 怎麼真正地
落地在教育當中
我覺得這是一個非常好的措施
因為從我的角度來說
你不能等到大學再去學 **AI**
我覺得 **AI** 應該像數學、語文
就是你自己的一個基礎的學科
因為你真正地
把它融入到自己的思想當中
未來可能有不同的專業的時候
你才自然而然的可以把它應用
所以我覺得是非常好的步驟
然後大學這個角度來說
比如說像香港理工大學
我們在大約 2020 年或 2021 年
學校就有一個人工智能的教育
給本科生的一個基礎的教育
就是大家都要去學
我覺得這也是非常重要的
那香港很多的大學

最近我們看到了很多的課程
就是「AI+」
比如說「AI+經濟」
「AI+ Materials
Science（AI+材料科學）」
我覺得是非常好的
就是在大學這裡充分地
開始學習 AI 怎麼能
真正地融入到未來的各行各業
等他們畢業的時候
自然而然一定會通過
自己的工具去實現
就是我剛才說的接下來的
AI native、AI 原生
最終的我們要到達的就叫
AI native economy
AI 原生經濟
最後請楊教授幫我們總結一下
給一些 AI 小錦囊大家
在日常生活中
我們怎樣用好 AI 或者學好 AI
我覺得大家一定要放心
大膽地去使用 AI
因為實話說，這一波的 AI
我們覺得我們真的是很幸運
也許這個技術實話說
如果沒有 OpenAI 的堅持
我覺得我們可能要等二、三十年
這是有可能的
但是因為他們堅持了不斷的
Scaling up（擴大規模）上去
GPT-4（生成型預訓練
變換模型 4）的 ChatGPT
這樣的一個突破
讓我們現在，尤其我覺得
國家最近的投入度是非常大的
我相信在這麼大的投入度
而且以及我們現在全民對 AI

都是真的非常興奮的一種情況下
一定會帶來重大的、非常多的
大家可能不能想像的一些突破
所以我的一個建議就是
大家放心大膽的在各行各業
把它用起來
感謝楊教授的分享
楊教授剛才提到多模態的 AI
現在會看、會聽和會理解
香港的優勢
就是在於專業和有深度
大膽去嘗試
就是我們擁抱這個 AI 年代的
最落實和最貼切的方向
這個節目《AI 大講堂》的目的
是透過一系列的專家對談
用深入淺出的方法
讓大家學會 AI 和善用 AI
希望大家繼續關注 AI 的發展
節目時間來到這裡，再見