

《AI 大講堂》

第四集 - AI 如何思考？它能夠如何走進市民日常生活？

視像文字稿

歡迎收看《AI 大講堂》
我是方健儀
越來越多人用人工智能
亦都會將 AI（人工智能）
變成我們的小工具
所以它們已經可以說
是我們的 24/7 小助手了
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
而這集的嘉賓
請來香港科技大學首席副校長
本身亦是香港生成式人工智能
研發中心（HKGAI）主任
郭毅可教授
郭教授近年是主導
香港大模型 HKGAI 的研發工作
他在節目裡會和我們講述
究竟 AI 是怎樣思考
和香港自主研發大模型的
重要性在哪裡
郭教授你好
你好
很多人都會用到 AI
他們覺得非常方便
它的便利程度或全面性
比一些互聯網搜尋器更好
不如和我們說說究竟
AI 的思考模式是如何運作？
我們過去在互聯網上
如果我們想知道一些問題
我們一般是去搜索內容
我們給個關鍵詞

然後搜索各種各樣的文章
自己去看，自己去總結
這是我們互聯網時代
那麼現在的 AI 不一樣
它是這樣的，它是一個
你要問問題，之後它去搜索
根據搜索的文章
發現相關的知識
然後把它總結出來告訴你
所以它實際上
開始你好像要在在一本書上
找到相關的幾頁內容
然後自己翻幾頁看
所以你要做兩個判斷
第一個，你收到的東西跟它有關
第二個，這內容裡面哪些
是要讀很多東西
然後得到一個回答
那麼今天這些工作 AI 做了
它可以直接地回答你
它理解你的問題
幫你找到相關的內容
最後總結出來回答你
這個基本上是這樣的
有很多知識它已經被訓練進去
所以它就可以直接的回答你
有些它不清楚的、不知道的
它可以去搜索網站
然後根據它對這些知識的理解
總結好再回答你
所以這個完全不一樣
所以你基本上
就好像有一個你的助手
幫你去獲得需要的知識再告訴你
所以這樣的話，效率是非常高的
聽你這樣說
其實 AI 是非常聰明的
但是它能否理解到人類的情感？

因為我們也有價值觀
也有道德邊界
AI 又能否領略得到？
這個就看我們怎樣訓練它
如果想到情感的問題
我給你舉個例子
我常常幹的事情
我拍一張照片
比如我幾天前去了大理
參加學生的婚禮
我就在大理的邊上拍了一張照片
照片上是我跟一個
我的學生新娘和新郎拍的照片
它就會給我寫一首非常好的詩
來祝賀新郎和新娘
你看這裡面你說 AI 懂不懂情感
它當然知道
首先它知道我為甚麼選大理
他就會寫到「蒼山洱海」
「海誓山盟」，它會寫出來
你說它不懂情感
那它怎麼會知道？
但是你說它也會知道
這是一個歡樂的場合
它就會寫到一些
非常美麗的語句來表達
所以你說「AI 有沒有情感」
和「AI 懂不懂得人類的情感」
是兩回事
我們主要是強調
「AI 懂得人類的情感」
那它為甚麼會懂得我們的情感
是我們訓練它
所謂的邊界也是一樣的
倫理也是一樣的
你要知道傷害別人是不好的
那你就要告訴它
所以在很多的場合

它是懂得人類的價值觀
和懂得人類的道德邊界
這是我們追求的
但是這不等於 AI 有情感
AI 有價值觀
AI 有邊界
我們現在追求的是
AI 能夠理解、能夠接受
能夠遵循我們的情感
我們的價值觀
我們的道德邊界
這是我們的工作
我們大概都知道分別在哪裡
其實 AI 的確是改變了
我們的日常生活
郭教授都提及過一個論調
人類和 AI 的共存
我們會有一個「人機共生」
這是一個甚麼概念？
「人機共生」的這個狀態
已經不是今天發生的
已經是開始很長一段時間了
那麼甚麼意思？
就是我們的智能和機器的智能
已經共存、互相依賴
現在比較厲害的是
無論甚麼事情
現在都離不開（人工智能）了
現在我們離不開一個導航
慢慢地你離不開
你的 chatbot（聊天機器人）
我希望你們，我們的努力方向
是希望我們的「港話通」
將來是香港市民的一個陪伴
對吧？
比如我們現在
我今天看了他們推出來的
就是我們交稅

香港今天的交稅
你怎麼交稅、怎麼填表
你一打它全告訴你
所以你就會覺得這個生活
已經慢慢跟一個機器的智能
連在一起了
這個已經是今天的一個常態
那麼未來為甚麼說它更加厲害
以後它是個 **Agent**（代理）
是智能體
英文叫做 **Agent**（代理）
你可以交代它做事情
例如我現在也常常做這種事情
我要做一件事情
比如希望後天我要去北京
那我就跟它說，我要去北京
我晚上 8 點才有時間
我希望你給我訂票
那以前我是讓我秘書做的
現在完全可以
讓 **Agent**（代理）去做
它給你搜尋網絡
給你找好票
然後你比較價格
然後去 **communicate**（溝通）
就這樣做的
所以這樣的話
你就發現這個「硅基智能」
就叫「碳硅共生」
或者「人機共生」，有各種說法
英文叫 **Symbiotic Society**
（共生社會）就是你跟它
完全生活在一起
第二個很重要的一點
你們還互相學習
我肯定要跟
我的 **chatbot**（聊天機器人）
學習很多東西

但反過來說你也教它
我發現它跟我不太一樣的話
我可以教它，讓它更像我
能夠幫我做點事情
更以我的方式
那麼這個也是一種學習，對吧？
所以它也在學你
這是我個人的
同時我們大模型
也不斷地學習今天新的知識
和我們對它的新要求
第三，所謂的「共生」
這是我們的社會秩序發生的變化
它是二元的
它有個機器，對吧？
它有智能，如果它幫你做事情
它做錯了誰負責？
所以很多我們的社會秩序
社會的一些常識
都發生了根本性的變化
所以我們要做很大的改變
未來的教育是甚麼樣子的
未來醫療是甚麼樣子的
未來整個的社會倫理
都會發生變化
那麼現在就有一個問題
我們怎麼來應付這個局面
你不能說你不許用人工智能
這個是不行、不通的
那你就改變你的作業方式
你要改變你的考試方式
這些就是「人機共生」以後
發生的巨大變化
所以人機共生這個概念
有很多不同的層次
而郭教授都提到
我們要去訓練 AI 的重要性
所以人們都不禁問

我們究竟怎樣可以訓練一個 AI？
訓練 AI 當然是一個技術活
但是你們可以把它想得比較簡單
就是跟訓練我們的腦子差不多
AI 的物理載體就是神經元網絡
實際上就是模擬
我們腦子裡怎麼工作
這是第一條
具體來說，我們的腦子怎麼運作
我們怎麼學習的
我們去看很多的、做很多的觀察
包括讀書
AI 也一樣，看很多的文字
首先這些文字要把它放到腦子
去組成這個知識的網絡
這個地方是通過一個編碼的過程
你看了過程以後
它就把你很多的語句
形成一種編碼
這個編碼有意思的地方就在於
它不僅是一個固定的編碼
而且這個編碼可以根據
你的語音、語境的環境
來做不同的改變，要計算
比如說一個蘋果
如果是個水果的話
它變成一個水果的編碼
如果是一個手機的話
它有手機的編碼
所以這個是很聰明的一套邏輯
這就彷彿在你的腦子中間
看到這個東西有一個神經元的連結
看到另外一個神經元的連結
這個時候就通過學習把它建立了
完成以後再做甚麼？
就看很多文字
我們叫下一個詞的預測
甚麼意思？

就是把整個語言的組成學會
學會以後機器就可以講話
當然可能這些話
沒有太大的邏輯性
第二步我們要把它
訓練成有邏輯的講話
你還要給它看各種領域的知識
它能夠講出很多話來了
這些話比如說
是關於這個社會的、新聞的
一堆語句連起來
還要再去看很多東西，學會這個
但這個還不夠
還要保證它講對話
以後還要給它、告訴它
這句話是不合道理的
這句話是不能講的
這句話是錯的
這句話是好的
這句話比這句話更好
應該這麼講
這個就要教它所謂的價值觀
或者叫「對齊」，像我們人類講話
這個講完以後就差不多了
那麼你就開始放下去用
用的過程才得到反饋
然後再去修正
我們先休息一下
待會回來繼續和郭教授聊天
休息回來
繼續《AI 大講堂》的時間
我們今天的嘉賓
是香港科技大學首席副校長
本身也是香港生成式人工智能
研發中心主任郭毅可教授
郭教授你好 你好
有些情況可能很多朋友都試過
就是我去跟它互動的時候

發現 AI 一本正經地說錯話
這個術語叫「AI 幻覺」
那個「AI 幻覺」為甚麼會出現
你自己有沒有經歷過？
這個我們叫「幻覺」
或者普通話叫
「一本正經地胡說八道」
在早期的 AI
它自己的一個語言系統學好以後
它的知識基本上
局限於你給它所有的例子
它沒有額外的知識進來
那麼在這個時候
它出錯的概率是很高的
因為它不知道一樣事情
但它又要告訴你一個答案
怎麼辦？只能編
因為早期的 AI
它還不能說「我不知道」
因為要告訴你不知道是很難的
因為它要搜遍它所有的東西
還發現組織不出來你的回答
所以怎麼辦？
它就告訴你一個似是而非的東西
這是過去
後來我們有很多的技術
就開始來彌補它
一個彌補的方法就是搜索
我發現，首先這已經開始
有非常強大的能力
我不是首先編出我的東西
我從已經學到的東西裡面告訴你
它是先去搜索網站
現在網上都得到相關的資訊
然後再用它過去的知識
來組織回答
譬如我舉個最簡單的例子
我訓練的一個模型

比如說用的是 2025 年的
可是我問了 2026 年的問題
我問昨天，它不可能知道
所以它只能瞎編
可是如果我現在是搜索
我問你昨天天氣是怎麼樣
它不是由大模式生成的
它是網上得到了昨天的信息
再把它組織成語言回答你
所以就沒有 hallucination
沒有幻覺
現在的幻覺率並不太高
因為它的網絡搜索能力
和我們做的判斷這些搜索的能力和修正它的能力
已經有了巨大的（能力）
它總是無止盡的
所以現在這個「AI 幻覺率」
聽郭教授說大大降低了
但是例如我作為一個用家
看到好像不是很準確的時候
可否輸入一個 Prompt（提示詞）
透過我給的指令
例如你可否給我一個答案時
你亦進行事實的核查
Fact-check（事實查核）
這樣會否做到將那個「AI 幻覺」
再進一步降低？
我覺得 AI 有個
literacy（素養）的問題
就是素養的問題
就像上課跟老師、學生對話一樣
老師的話你也不可以全信
所以要有一個批判性思維
所以作為 AI 的一個素養
實際上我們對培養
Critical Thinking
（批判性思維）變得很重要

就是你對一些事情
你要問一下
特別是重要的事實
你可以信它百分之九十五以上
但是我覺得你最好
對你很關鍵的問題可以查一下
更重要的是
你可以繼續詢問，對吧
你覺得這個答案你理解了
你可以再問它為甚麼
你怎麼獲得的
而且現在很多的 AI 系統
包括像「港話通」
它都會告訴你，它的參考
特別是事實，你自己要去查
所以這些都是今天這個事情
所以我覺得如何使得 AI
給你的支持更為可信，是兩條
第一個，我們的機器
要 **generate**（生成）
要有技術來保證問題的可信性
第二我要告訴你讓它透明一點
我告訴你這個知識從哪裡來的
第三點是你自己
你要覺得這是 AI
所以其實這對人類的智力的提高
是非常有利的
現在的 AI 就是這樣的狀態
所以大家都要做守門員把關角色
都非常重要
對，你自己把關
郭教授，聊一下你現在
主導的 HKGAI
就是現在本地自主研發的大模型
你也有提過這個 HKGAI
是要「很香港」的
即是甚麼意思？
就是「識香港事

說香港話
為香港人做事」
所以就變成了有一個本地的語境
但是很多人都說廣東話很難學
所以你去做一個香港語境的時候
第一，你有沒有遇到困難？
第二，都知道 HKGAI 不斷有更新
現在的版本
已經可以符合市民的要求
或者工作效率提升
但是有沒有一些例子可以說
最初的時候我們發現到一些問題
然後去到現在這個版本
已經可以彌補
或者已經可以解決
可以幫助到市民？
我覺得第一
為香港人民做一個自己的模型
做一個自己的
chatbot（聊天機器人）
做一個懂香港人
像香港人
為香港人
香港人民的好夥伴
這樣一個 **Agent**（代理）
這是一個非常有意義的事情
也非常的有挑戰性
難點在幾個地方
第一，香港是個國際化的社會
它有自己的法律系統
有自己的治理規則
我們有一國兩制的特點
所以我要反映香港這個價值
這個就有意思
你怎樣令它的行為
像一個知法、懂法、守法的香港人
這是第一步
這個要放在模型裡面

這很有意思
所以這個是我們做訓練的時候
最大的挑戰
如果把這個挑戰做好了以後
那你就有可能對於
任何一個社會系統和社會環境
你可以把一個模型
牽到這個社會環境上去
所以這對中國的模型的出海
是非常有用的
還有一個東西就是你的標註
這些都是很大的工程，你怎麼做好
那麼這樣的話它就懂了
還有很多非常有意思的
香港有香港人民的生活方式
比如說香港人民對於買便宜貨
非常有興趣
對於出去出遊的時候
怎樣的交通有興趣
常常問在這個地方
甚麼地方有最近的咖啡店
這些問題
這個問題其實很容易出問題
你說「請你告訴我
立法會附近的咖啡店」
甚麼叫附近？
200 米？1,000 米？
都是附近
它說的是我能夠走過去的
我過得了馬路的咖啡店是不是？
這個機器要求的推論就很有趣
所以這一些都是一個
比如說 Gemini 這種模型
或者都做不到
因為它不知道香港
這些我們都花努力去做的
而且「港話通」很大的一個特點
我們要做的是香港生活的服務

比如我跟你講的
今天的交稅
我不能錯
我出錯了以後它交稅
那就大問題了
所以我們要花很大的努力
把這些做好
做好這些就跟訓練
就跟我們教孩子是一樣的
你要用很好的數據
很好的標註
所以我們要與時並進
最後請教郭教授
能否給一些 AI 小錦囊給觀眾？
就是我們在日常生活中
怎樣可以用好 AI？
第一個小錦囊
下載「港話通」
第二個小錦囊
問問題的時候
我覺得你問問題很講究
所以我們會馬上推出來一個應用
叫做「好問題」
這個是甚麼意思？
我們交流交流
你也有甚麼好問題，大家分享
一個問題問得好和問得不好
你獲得的東西完全不一樣
AI is question
sensitive（對問題敏感）
你問得好
你給的 **information**（資訊）多
hint（提示）多
它就給你好回答
你問的 **too general**（太普通）
它就給你普通的回答
那麼我們怎樣去學習好問題
一個是幫助我們

一起進化「港話通」
第二點就是我們共同演進
這個是需要訓練它
訓練它，不是我們訓練
最好是所有的香港人民
一起訓練它
那我們就把它做得很好
所以希望這個錦囊的最關鍵
就是你們能夠反饋給我們
我們爭取把這個東西
做成大家真正喜歡的東西
多謝郭教授的分享
郭教授剛才也跟我們說了
AI 究竟是如何思考
以及我們自主研發的大模型
HKGAI 的重要性
最重要的是大家一起訓練
AI 就會成為我們的好助手
這個節目《AI 大講堂》
是希望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
歡迎大家繼續關注
節目時間來到這裡，再見