

<https://www.wsj.com/articles/will-ai-help-us-or-leave-us-behind-a1dfa5d3>

文章

人工智慧會幫助我們還是把我們拋在後面？

在我們眼前展開的革命最緊迫的問題是，我們能否將技術與我們的人性聯繫起來。

沃爾特·以撒森報導

美國東部時間 11 年 2023 月 9 日 上午 59 : <>

技術革命通常始於很少大張旗鼓。1760 年的某個早晨，沒有人醒來大喊：“天哪，工業革命才剛剛開始！即使數位革命也在幕後消失了很多年，業餘愛好者將個人電腦拼湊在一起，在諸如自製計算機俱樂部等令人討厭的聚會上炫耀，然後人們才注意到世界正在發生根本性的變化。

人工智慧革命是不同的。在 2023 年春季的幾周內，數百萬具有技術意識的人和普通人注意到，一場變革正在以驚人的速度發生，這將改變工作、學習和創造力的性質以及日常生活的任務。

機器自己變得超級智慧是不可避免的嗎？

聊天機器人和其他形式的生成人工智慧的出現——可以通過在大量數據集上進行自我訓練來生成原始文本或圖像的計算機——提出了一個對人工智慧歷史至關重要的問題。我們的目標應該是讓人類和機器緊密地成為合作夥伴，以便它們通過共生工作取得進步嗎？還是機器將不可避免地自行變得超級智慧，最終將我們這些凡人拋在後面？

第一所學校的守護神是 19 世紀的英國作家和數學家阿達·洛夫萊斯。在 1840 年代，她描述了她的朋友查理斯·巴貝奇（Charles Babbage）正在建造的數字計算機中使用穿孔卡如何使機器不僅可以處理數位，還可以處理任何可以用符號記錄的東西，包括文字、音樂和圖像。換句話說，她設想了一種通用計算機。她說，它甚至能夠執行「抽象的心理過程」。但她補充了

一個重要的警告：與人類不同，它缺乏創造力或自己提出想法的能力。“它可以做任何我們知道如何命令它執行的事情，”她寫道，但它“沒有任何自命不凡的東西。

一個世紀後，英國數學家和計算機科學家艾倫·圖靈成為第二學派的守護神。他在1950年發表的關於人工智慧的開創性論文《計算機與智慧》(Computing Machinery and Intelligence)提出了一個問題：“機器能思考嗎？在一個名為「洛夫萊斯夫人的反對意見」的一節中，他反駁了她的論點，他說有一天我們會開發出能夠自己學習的機器，而不僅僅是執行程式。他們將能夠提出想法。“與其嘗試製作一個類比成人思維的程式，為什麼不嘗試製作一個類比孩子思維的程式呢？”他問道。“如果這受到適當的教育課程，人們將獲得成人的大腦。

他提出了一種著名的「模仿遊戲」，在這種遊戲中，計算機將嘗試回答問題並進行對話，其技巧使人無法猜測它是機器還是人類。隨著OpenAI的GPT-4和谷歌的Bard在三月份的發佈，很明顯現在有機器可以輕鬆通過這個圖靈測試。

到目前為止，Lovelace試圖在人類創造力和機器處理能力之間建立更緊密的夥伴關係 - 增強智慧而不是人工智慧 - 已經佔了上風。數位革命是由人機介面的改進推動的，人機介面是我們用來盡可能無縫和快速地與機器交換資訊的方法。

這種方法的先驅是麻省理工學院的心理學家和技術夢想家J.C.R. Licklider。在研究人類可以立即與防空計算機交互的方法時，他幫助設計了使用視頻屏幕來輸出資訊併為人類提供輸入回應的方法。這一經歷為現代技術史上最具影響力的論文之一奠定了基礎，題為“人機共生”，Licklider於1960年發表。“希望是，用不了多少年，人類的大腦和計算機將非常緊密地結合在一起，”他寫道，“由此產生的夥伴關係將像人類大腦從未以我們今天所知道的資訊處理機器所沒有的方式思考和處理數據一樣思考和處理數據。

麻省理工學院的學生將其中一些早期計算機變成了帶有操縱桿和指標的視頻遊戲。連接機器和人類的其他進步很快隨之而來，例如我們現在在個人計算機上使用的點擊滑鼠和圖形用戶介面。在2011年蘋果的最後一次會議上，史蒂夫·賈伯斯（Steve Jobs）看到並測試了Siri，這是下一個偉大進步的一個例子：使用語音與我們的機器進行交互的能力。

埃隆·馬斯克（Elon Musk）現在正在追求人機思維融合的最終一步：一種帶有神經感測器的微晶元，可以植人大腦，並允許我們和計算機之間幾乎即時共用資訊和信號。本月，他的公司Neuralink完成了最後一輪動物研究，並準備向美國食品和藥物管理局（Food and Drug Administration）申請，允許將晶元植入人體測試對象的大腦中。“確保人工智慧不會反對我們或摧毀人類的最佳方法是將其與人類代理緊密聯繫起來，”他說。

人工智慧的最新進展可能令人不安的是，這些系統遵循圖靈的願景，正在被設計成可以自己學習。他們甚至似乎正在發展自己的能動性，而不是被人類的指示和意圖所束縛。最可怕的是，它們的創造者，就像弗蘭肯斯坦博士一樣，有時並不完全理解他們如何以及為什麼產生某些答案，產生幻覺，做出錯誤的斷言，甚至似乎表達了令人不安的情緒。這引發了即將到來的「奇點」的幽靈，數學家約翰·馮·諾依曼（John von Neumann）和科幻作家弗諾·文格（Vernor Vinge）用這個詞來描述人工智慧可以以無法控制的速度自行前進並將我們人類拋在後面的時刻。

人工智慧安全將不僅僅來自召集眉頭緊鎖的人來制定國際法規或人工智慧行為準則，儘管這些努力可能是有價值的。相反，我認為這將要求我們回到尋找新的更好的技術的方法，將人類和計算機更緊密地聯繫在一起。目標應該是確保我們的機器始終與人類代理機構相連。至少，這將確保真正的人類對機器的行為負責。充其量，這將使這些系統不太可能橫行霸道並發展自己的思想。

這個學期，在我在杜蘭大學教授的數字技術歷史課上，我和我的學生分享了看到一場新革命突然誕生的興奮。我讓他們讀了威廉·華茲華斯（William Wordsworth）的詩《法國大革命在開始時出現在狂熱者面

前》,其中有令人振奮的詩句：“在那個黎明中活著真是太幸福了！但我也指出，這個標題有一種前衛的諷刺意味。法國大革命的結局並不好。在這個黎明中成為狂熱者是可以的，但專注於保持革命與我們的人性聯繫也很重要。

以撒森正在寫一本埃隆·馬斯克 (*Elon Musk*) 的傳記。