

技術詞彙表

本技術詞彙表載有對本文件中使用的若干技術詞彙的解釋。因此，該等詞彙及其含義未必與該等詞彙的業內標準含義或用途一致。

「AI」或「人工智能」	指	人工智能
「AI智能體」	指	一個由 Agentic 人工智能技術驅動的系統，能夠自主感知環境、處理信息並執行動作，以實現特定目標，可以為軟件形式或嵌入硬件中
「Agentic 人工智能」	指	超越內容生成範疇，具備自主決策、執行動作及動態環境適應能力，以自主行動為核心，在極少人工干預下實現特定目標
「API」	指	應用程序接口，一種方便不同計算機系統之間交換信息及執行指令的計算機編程方法
「CAGR」	指	複合年增長率
「CDC」	指	變更數據捕獲，一種識別及跟蹤數據庫中數據變更的軟件過程
「數據融合」	指	通過整合多數據源，生成比單一數據源更一致、更準確且更有用的信息的過程
「數據治理」	指	管理企業系統中數據的可獲得性、可用性、完整性及安全性的過程
「數據湖倉」	指	一種現代數據體系結構，通過結合數據湖（原始形式的大型原始數據存儲庫）及數據湖倉（結構化數據的組織集）的關鍵功能來創建單一平台
「蒸餾」	指	將知識從大模型轉移到小模型的過程
「ETL」	指	提取－轉換－加載，指將數據從數據源中提取、經轉換後加載到目標系統的通用流程
「生成式人工智能」	指	通過從海量數據中學習並運用習得知識生成全新原創性內容，旨在覆蓋文本、圖像、代碼、音頻乃至視頻的創作，其生成結果可模擬人類創造性思維特徵

技術詞彙表

「混合檢索」	指	將不同的搜索索引和查詢策略組合起來，以確定給定查詢的最相關資料的過程
「大模型」	指	一種包含超大規模參數（通常為數十億或更多）的神經網絡架構，使其能夠處理跨大量數據集的複雜模式及關係
「微調」	指	基於規模較小的特定數據集對預訓練模型進行進一步訓練，以實現對具體應用的定制化適配的過程
「幻覺」	指	在人工智能的語境中，指人工智能模型產生的不正確或具誤導性的結果，這可能是由多種因素造成，包括訓練數據不足、模型作出的錯誤假設或用於訓練模型的數據存在偏差
「Iceberg」	指	用於分析SQL表的開源項目，專為高性能及易用性而設計
「IT」	指	信息技術
「KV緩存管理」	指	一種內存優化策略，在自回歸推理過程中選擇性地緩存及重用來自先前序列token的鍵值對
「千瓦時」	指	千瓦時，一種電功率量度單位，表示一件功率為一千瓦的設備在使用一小時之後所消耗的能量
「數學模型」	指	使用數學概念及語言對具體系統的抽象描述
「MCP」	指	模型上下文協議，一個開放的協議，可以在數據來源及AI驅動的工具之間實現安全的雙向連接
「MQL」	指	指標查詢語言，我們的專有技術，可實現複雜的數據資產檢索
「多模態數據」	指	以多種不同格式捕獲的數據，例如文本、圖像、音頻、視頻及其他形式的感官輸入

技術詞彙表

「開源」	指	可免費獲取、自由修改及重新分發的源代碼
「問答」	指	問答
「研發」	指	研究與開發
「RAG」	指	檢索增強生成，一種使生成人工智能模型能夠檢索及納入新信息的技術
「強化學習」	指	一種基於獎懲機制在動態環境中對預訓練大模型進行行為優化的適配技術
「SFT」	指	監督微調，一種用於使用包含深度領域特定知識的自定義標記數據集，使預訓練的大模型適應特定下游任務的技術
「SQL」	指	用於編程的領域特定語言，專為關係型數據庫管理系統中的數據管理，或關係型數據流管理系統中的流處理而設計
「半結構化數據」	指	具有某些確定或一致的特徵但不符合剛性結構的數據
「平方米」	指	平方米
「結構化數據」	指	使用預先定義和預期格式的數據
「分詞」	指	將廣泛的原始複雜數據(可能包括文本數據、圖像、文檔及公式)轉換為大模型可以理解及處理的格式的過程，同時保留數據中的語義關係及上下文細微差別，為大模型訓練奠定基礎
「噸」	指	公噸，重量單位，一公噸等於1,000公斤或2,204.6磅

技術詞彙表

「非結構化數據」	指	並無預先定義格式或組織的數據 (例如文檔、圖像及視頻的形式，通常涉及複雜的行業特定知識)，包括大部分企業數據，需要高級數據功能來管理和處理
「VL」	指	視覺語言模型，一種融合了計算機視覺及自然語言處理能力的AI模型