

## 技術詞彙表

本技術術語表包含對本文件中使用的若干技術術語的解釋。因此，該等術語及其涵義可能與該等術語的標準行業涵義或用法不一致。

「AAAI」	指	人工智能促進協會，成立於1979年(最初名為「美國人工智能協會」)，是一個國際性非營利科學協會，致力於促進人工智能的研究、教育和負責任的使用。AAAI每年舉辦AAAI人工智能大會－人工智能領域的頂級學術會議之一－旨在促進全球人工智能理論和應用領域的研究人員、從業人員和決策者之間的交流。
「ACL」	指	計算語言學協會，為一個專注於計算語言學和自然語言處理的頂級國際科學和專業協會。其旗艦會議(ACL年會)是NLP研究領域最頂級的同行評審會議之一。
「AI智能體」或「智能體」	指	一種軟件系統，旨在通過與環境互動及作出決策，自主執行任務以實現特定目標。
「API」	指	應用程序編程接口，允許不同軟件應用程序相互溝通的一組規則、協議及工具。
「標桿客戶」	指	2023年、2024年及2025年，貢獻總收入超過人民幣3.0百萬元的客戶或未達到此閾值但被視為對我們的業務具有戰略重要性的客戶。
「跨模態」	指	不同模態之間的互動或關係，着重於如何將一種模態的信息與另一種模態的信息進行映射、轉譯或與之對齊。
「註釋」	指	用相關信息或元數據對原始數據(如文本、圖像、音頻或視頻)進行標注或標記，使其易於理解並可用於機器學習模型的過程。此為監督學習的關鍵步驟，其中模型需要標注數據集來學習規律並做出準確預測。
「數據治理」	指	確保組織數據資產妥善管理的框架、政策、流程和標準。其側重於定義角色、責任和規則，以確保數據質量、安全性、合規性以及在整個組織內有效使用數據。

---

## 技術詞彙表

---

「數據管理」	指	在數據的整個生命周期內，用於有效收集、存儲、組織、維護和利用數據的操作流程、工具和技術。其重點是確保數據的可訪問性、可靠性，並可用於決策及運營。
「數據集」	指	為結構化數據的集合，被組織成有意義的格式，通常用於機器學習算法的分析、訓練或處理。
「決策智能」	指	設計、整合和協調先進的人工智能技術，以系統化地強化、自動化和解釋複雜的決策過程。
「數字化轉型」	指	將先進數字化技術及數據驅動策略融入組織運營的根本性過程，旨在利用數字化工具促進創新、提升效率，重塑營運模式，提升客戶數據價值，並加強競爭優勢。
「EMNLP」	指	自然語言處理實證方法會議，為自然語言處理和機器學習研究領域的頂級年度會議，自1996年以來一直由ACL及其SIGDAT小組主辦，被公認為全球三大高影響力NLP會議之一。
「端到端」	指	描述一個系統或流程，能夠在單一整合的工作流程中處理從初始原始輸入到最終輸出的完整任務，無需人工干預或單獨的中間階段。在我們的AI服務中，這通常指我們利用單一、多功能的大模型執行一系列複雜任務的方法，否則該等任務將需要多個專用模型組成的流程。
「異構數據」	指	由各種類型、格式或結構組成的數據，通常來自多個來源。其包含結構化、半結構化及非結構化數據，例如將數據庫中的數值數據與文件中的文字以及媒體庫中的圖像結合。
「知識圖譜」	指	以實體、其屬性及其實體之間關係的形式體現知識，使機器能根據上下文理解及推斷資料的結構化的架構。其增強語義理解，促進來自不同來源的數據整合，並提高搜尋的相關性，使其在搜索引擎、人工智能及數據分析等應用程序中顯得至關重要。

---

## 技術詞彙表

---

「LLM」	指	大型語言模型，是在超大文本數據集上通過自我監督機器學習訓練出來的一類基礎人工智能模型，能夠理解和生成人類語言以用於各種自然語言處理任務，尤其是文本生成。
「多模態」	指	AI系統在單一統一的框架內同時處理、理解及推理來自多種不同數據類型（模態，如文本、圖像、音頻及視頻）的信息的能力。
「運算子」	指	一種AI系統或工作流程中執行特定預先定義功能的基本建構塊。例如，數據攝取運算子、數據轉換運算子或模型推理運算子，該等運算子可以組合起來建立複雜的AI應用程序。
「預先訓練」	指	訓練機器學習模型的初步階段，特別是在自然語言處理(NLP)及深度學習中。在此階段中，模型在大數據集上進行訓練，以學習數據的一般模式、表現及特徵，然後再針對特定任務進行微調。
「結構化數據」	指	以預先界定的格式組織的信息，通常以關係數據庫或電子表格中的行及列為單位，使其容易搜尋及分析。其遵守嚴格的模式或數據模型，例如具有明確界定欄位的數值、類別或文字值。
「監督微調」	指	機器學習的一個流程，使用標記數據進一步訓練預先訓練好的模型，使其專門用於特定任務。此方法可確保模型從標記資料中學習模式及關係，以改善其在所需任務上的效能。
「TMM」	指	測試成熟度模型，為一個結構化框架，最初由伊利諾伊理工學院開發，用於評估和改進企業軟件測試流程的成熟度。
「非結構化數據」	指	無預設格式或組織結構的信息，因此很難使用傳統的關係數據庫進行儲存及分析。其需要專門的工具及技術，例如自然語言處理(NLP)或機器學習，才能進行有效的分析。