

Media Features

Hong Kong Economic Journal | 20 April 2026

Interview with Stephen Tsang, Chief Revenue Officer, NTT Com Aia; and Steven So, Chief Operating Officer, NTT Com Asia

NTT 擬下半年在港推 GPU 服務

人工智能 (AI) 正重塑各行各業，日本電訊巨頭 NTT 集團旗下香港子公司 NTT Com Asia 首席營運總監蘇耀宗【圖】預期，隨着大模型技術成熟，金融業 AI 應用將呈現爆發性增長。NTT 今年初啟動「金融 AI 架構」(Financial AI Fabric) 策略，建基於去年 11 月推出的超低時延 APN (All-Photonic Network, 全光網絡) 服務，計劃今年下半年在香港推出 GPU 服務 (GPU as a Service)，回應市場對 AI 專用算力的需求，並有意打造金融 AI 應用生態平台，匯聚生態夥伴的 AI 應用，支撐未來金融 AI 發展。

NTT 擬下半年在港推 GPU 服務

人工智能 (AI) 正重塑各行各業，日本電訊巨頭 NTT 集團旗下香港子公司 NTT Com Asia 首席營運總監蘇耀宗【圖】預期，隨着大模型技術成熟，金融業 AI 應用將呈現爆發性增長。NTT 今年初啟動「金融 AI 架構」(Financial AI Fabric) 策略，建基於去年 11 月推出的超低時延 APN (All-Photonic Network, 全光網絡) 服務，計劃今年下半年在香港推出 GPU 服務 (GPU as a Service)，回應市場對 AI 專用算力的需求，並有意打造金融 AI 應用生態平台，匯聚生態夥伴的 AI 應用，支撐未來金融 AI 發展。



服務金融客戶。蘇耀宗透露，目前 NTT 在香港有超過 300 個金融客戶，有必要為未來金融 AI 應用的爆發性增長做好準備，「過去兩年只是一個起步，當 LLM (大語言模型) 技術成熟，我們認為好多金融應用會建立在大語言模型之上」。他認為，DeepSeek 推動開源模型發展，將來不少 LLM 會部署在私有環境中，尤其在金融業。由於受到高度監管，相信在私有環境開發金融 AI 應用是大勢所趨。

去年 11 月，NTT 在香港率先推出「APN InterLink」服務，採用全光網絡技術。傳統網絡的接收器之間涉及光電轉換，導致延時和能源流失；而 APN 的兩台接收器中間基本只有一條光纖，省卻多餘光電轉換環節，進而大幅減低延時和節省能源。數據顯示，APN 連接可把延時降低 30%，能源效率提升 40%。

蘇耀宗說，在 AI 時代，APN 技術大有用處，例如利用高速網絡統籌多地的算力資源、同步調用部署在不同地方的 AI 模型等，為客戶提供更多靈活性。

NTT Com Asia 首席營收總監曾慶德表示，APN 技術已做好商業化準備，正值 AI 發展推升此類服務需求，未來金融 AI 應用若能提供成本效益，有信心客戶會積極採用。

作為 APN 服務的延伸，NTT 擬今年下半年推出 GPU as a Service，主要因為 AI 應用，特別是金融業模型訓練與推理需求明顯上升，帶動 AI 算力需求。蘇耀宗指出，除了自己的 GPU，NTT 將與在香港擁有 GPU 資源的公司合作。在應用層面，NTT 則會建造生態平台，匯聚生態夥伴的 AI 應用，並引入私有 AI 環境中運行，在金融級的合規、安全及數據治理框架下，推動 AI 落地。

採訪、撰文：陳健輝

打造金融 AI 應用生態平台

由於旗下金融數據中心鄰近港交所 (00388) 數據中心，NTT 一直主力服務金融客戶。蘇耀宗透露，目前 NTT 在香港有超過 300 個金融客戶，有必要為未來金融 AI 應用的爆發性增長做好準備，「過去兩年只是一個起步，當 LLM (大語言模型) 技術成熟，我們認為好多金融應用會建立在大語言模型之上」。他認為，DeepSeek 推動開源模型發展，將來不少 LLM 會部署在私有環境中，尤其在金融業。由於受到高度監管，相信在私有環境開發金融 AI 應用是大勢所趨。

打造金融 AI 應用生態平台

由於旗下金融數據中心鄰近港交所 (00388) 數據中心，NTT 一直主力服務金融客戶。蘇耀宗透露，目前 NTT 在香港有超過 300 個金融客戶，有必要為未來金融 AI 應用的爆發性增長做好準備，「過去兩年只是一個起步，當 LLM (大語言模型) 技術成熟，我們認為好多金融應用會建立在大語言模型之上」。他認為，DeepSeek 推動開源模型發展，將來不少 LLM 會部署在私有環境中，尤其在金融業。由於受到高度監管，相信在私有環境開發金融 AI 應用是大勢所趨。

去年 11 月，NTT 在香港率先推出「APN InterLink」服務，採用全光網絡技術。傳統網絡的接收器之間涉及光電轉換，導致延時和能源流失；而 APN 的兩台接收器中間基本只有一條光纖，省卻多餘光電轉換環節，進而大幅減低延時和節省能源。數據顯示，APN 連接可把延時降低 30%，能源效率提升 40%。蘇耀宗說，在 AI 時代，APN 技術大有用處，例如利用高速網絡統籌多地的算力資源、同步調用部署在不同地方的 AI 模型等，為客戶提供更多靈活性。

NTT Com Asia 首席營收總監曾慶德表示，APN 技術已做好商業化準備，正值 AI 發展推升此類服務需求，未來金融 AI 應用若能提供成本效益，有信心客戶會積極採用。

作為 APN 服務的延伸，NTT 擬今年下半年推出 GPU as a Service，主要因為 AI 應用，特別是金融業模型訓練與推理需求明顯上升，帶動 AI 算力需求。蘇耀宗指出，除了自己的 GPU，NTT 將與在香港擁有 GPU 資源的公司合作。在應用層面，NTT 則會建造生態平台，匯聚生態夥伴的 AI 應用，並引入私有 AI 環境中運行，在金融級的合規、安全及數據治理框架下，推動 AI 落地。