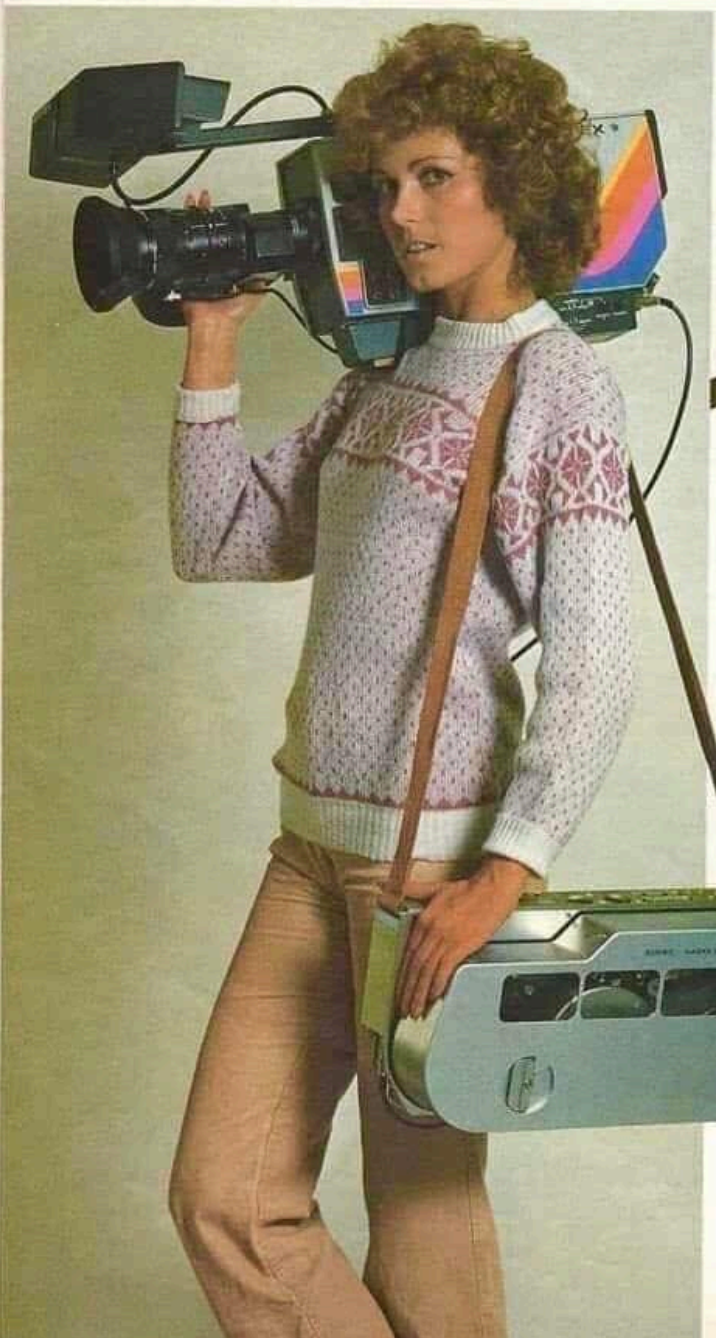


# NAB Show Las Vegas and Google Next

## 2025 出差報告

A NEW PORTABLE "C" VTR  
THAT WEIGHS ONLY 15 POUNDS.

THE  
AMAZING  
VPR-5.



We started a revolution by combining our talents with those of Kudelski SA, an innovator in audio technology, to create the Ampex Nagra VPR-5.

At 15 pounds (including 20-min. reels, battery and cover), the VPR-5 is the world's lightest and most portable one-inch Type "C" VTR.

VPR-5 is as small as a briefcase, measuring  $44.2 \times 21.5 \times 13.7$  cm, yet it delivers the quality and editing capability demanded from any full performance C-format VTR.

VPR-5 features: dual micro-processors, two high-quality audio channels plus a SMPTE/EBU time code channel, dual-cue editing, LCD display of VU and PPM levels, audio confidence playback selector independent of video, and a rugged aluminum alloy casing.

Get all details from your nearest Ampex International sales office.

**AMPEX**  
Ampex Corporation • One of The Signal Companies  
Europe, Africa, Middle East H.Q.,  
Ampex International,  
Acre Road, Reading, England.  
Tel: 0734. 875200 Telex: 847611

# 前言：

NAB Show 與 Google Cloud Next 這兩個活動剛好接連著舉行，因此形成一個特殊的機會，可以一次看完一個特定產業的前世、今生與未來。本次出差旨在深入了解全球串流媒體產業的最新趨勢與技術發展，特別針對播送工程(distribution engineering)與內容製作(content production)兩大核心領域進行考察與學習。透過參與在拉斯維加斯舉行的 NAB Streaming Summit 2025以及 Google Cloud Next 2025, 我得以與業界領袖、技術專家進行交流，並親身體驗最新的解決方案與創新應用。本報告將針對這兩個活動的觀察與心得進行彙整，並分析其對本會「公視+」OTT 串流平台未來發展的潛在影響與啟示。

## 1. 在播送工程方面，Streaming Summit 的重點議題與技術趨勢包括：

- a. **低延遲串流技術(Low-Latency Streaming)**：隨著互動式串流應用(如體育賽事、線上競標、遊戲直播等)的需求日益增長，低延遲成為了業界關注的焦點。多家技術供應商展示了最新的低延遲串流解決方案，例如基於 WebRTC 的增強方案、優化的傳輸協議(如 SRT、RIST)以及邊緣運算技術的應用。這些技術旨在將點到點的延遲降低至毫秒級別，提升使用者觀看體驗的即時性與互動性。
- b. **內容分發網路(CDN)的智慧化與邊緣化**：CDN 在確保串流內容穩定、高效傳輸方面扮演著至關重要的角色。會議強調了 CDN 的智慧化發展趨勢，例如基於 AI/ML 的流量預測與負載均衡、動態內容加速、以及更精細化的地理位置路由。同時，邊緣運算的興起使得將計算與儲存資源更靠近使用者成為可能，進一步降低延遲、提升傳輸效率。

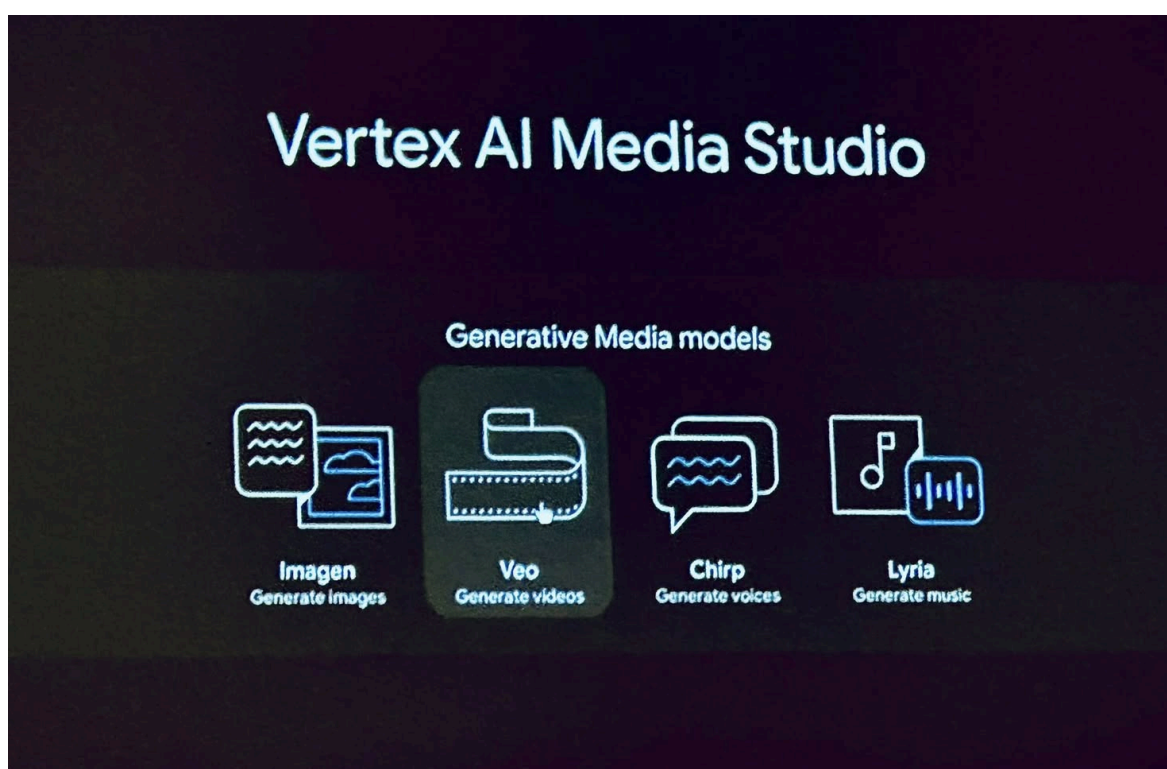
## 2. 在內容製作方面，Streaming Summit 的焦點主要集中在雲端化製作流程、AI 輔助內容生產以及高品質內容的規模化生產：

- a. **雲端製作工作流程(Cloud-Based Production Workflows)**：越來越多的內容製作工具與平台遷移至雲端，實現了異地協作、彈性擴展以及降低硬體成本的優勢。本次會議展示了包括雲端剪輯、雲端調色、雲端特效製作以及雲端資產管理等多種解決方案。
- b. **AI 輔助內容生產(AI-Assisted Content Production)**：人工智慧技術正在滲透到內容產製的各個環節，例如自動化字幕生成、影片內容分析、智慧剪輯、甚至部分創意內容的生成。本次會議展示了多種基於 AI 的工具與平台，旨在提升內容生產的效率與品質。
- c. **高品質內容的規模化生產(Scalable High-Quality Content Production)**：在競爭激烈的 OTT 市場中，持續提供高品質的原創內容是吸引和留住使用者的關鍵。本次會議探討了如何利用技術和流程創新，實現高品質內容的規模化生產，例如虛擬製作技術、自動化工作流程以及協作平台。

## 3. GCP 在播送工程方面提供的關鍵技術與服務包括：

- a. **全球基礎設施與 CDN 整合**：遍布全球的資料中心網路，可提供高可用性、低延遲的雲端運算與儲存資源。同時，CDN 供應商進行了深度整合，方便 OTT 平台構建全球性的內容分發網路。
- b. **彈性擴展與自動化**：雲端服務具有高度的彈性擴展能力，可以根據流量需求自動調整資源，應對突發流量高峰。同時，自動化工具，可簡化部署、管理與監控流程。

- c. 數據分析與監控：GCP 提供了強大的數據分析與監控工具，可以幫助 OTT 平台實時監控服務狀態、分析使用者行為、優化內容推薦以及預測潛在問題。
4. GCP 在內容製作方面提供的關鍵技術與服務包括：
- a. 雲端儲存與資產管理：安全、可靠、可擴展的雲端儲存服務，方便 OTT 平台存儲和管理大量的影音內容資產。同時，智慧化的資產管理工具，方便內容的檢索、組織和協作。
  - b. **AI 與機器學習服務**：豐富的 AI 與機器學習服務，例如影片分析、語音轉文字、自然語言處理等，可以應用於內容生產的各個環節，例如自動化字幕生成、內容標註、情感分析等。
  - c. 協作與工作流程平台：一系列的協作與工作流程平台，例如 Google Workspace，便於內容製作團隊進行溝通、協作和項目管理；而首次發表的創作者工具 Vertex AI Media Studio 加速實現 AI 內容共創，提高創作效率，更進一步降低內容製作的門檻與製作成本。



總體來說，已經可以看到電視產業，在播送工程與內容製作兩大面向，有全新的挑戰與契機。播送工程面的低延遲串流、AV1 編碼、智慧化 CDN 以及邊緣運算等技術將持續演進，為提升使用者體驗和降低運營成本帶來新的機遇。內容製作的雲端化、AI 輔助以及規模化生產將成為關鍵趨勢，推動內容生產效率和品質的提升。

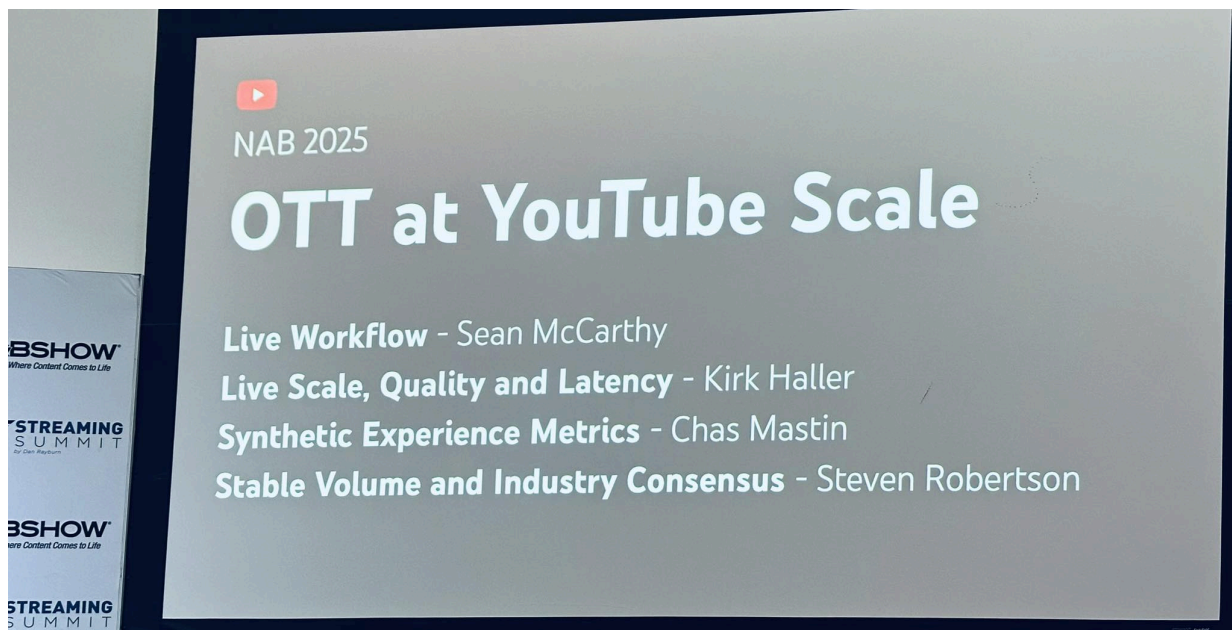
以下將透過兩個案例，YouTube 與 TikTok，如何以串流優化策略，來提升觀眾滿意度。最後介紹 Google 推出的 Dubbing AI 方案，將如何對影視產業造成革命性影響。

備註：本報告撰寫時，使用 AI 工具：ChatGPT, Gemini Pro.2.5, Gemini Deep Research



# 第一部分

## 第一章 OTT 平台技術報告: YouTube 的實踐與創新



### 1. 引言

隨著網際網路技術的飛速發展，OTT(Over-The-Top)串流媒體服務已經成為主流的影片內容分發方式。YouTube 作為全球領先的影片平台，在建構和營運大規模 OTT 服務方面累積了豐富的經驗。這個段落，將深入分析 YouTube 在 OTT 平台建設中遇到的技術挑戰，並詳細闡述其創新性的解決方案，以作為 OTT 產業的技術發展提供參考。

### 2. YouTube OTT 平台概覽

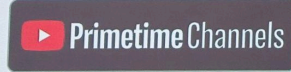
YouTube 在台灣的功能，大家熟悉的主要是在「社群互動影音平台」的領域，也就是說，可以讓觀眾一邊觀看一邊在聊天室裡互動，還可以一邊抖內(如超級留言或付費會員)。但在美國市場，YouTube 的功能並不只有社群互動影音，OTT 業務更是一大重點，且涵蓋多種產品和服務，包括：

- **YouTube TV**: 提供直播線性電視節目的 OTT 服務，類似於傳統的有線電視服務。
- **Sunday Ticket**: 作為 YouTube TV 的附加服務，以及獨立銷售的產品，提供體育賽事的直播。
- **Prime Time Channels**: 提供 24/7 全天候線性頻道的服務。



## OTT Purpose

Provide a personalized destination for transformative media experiences, connecting users with their passions on a global scale.

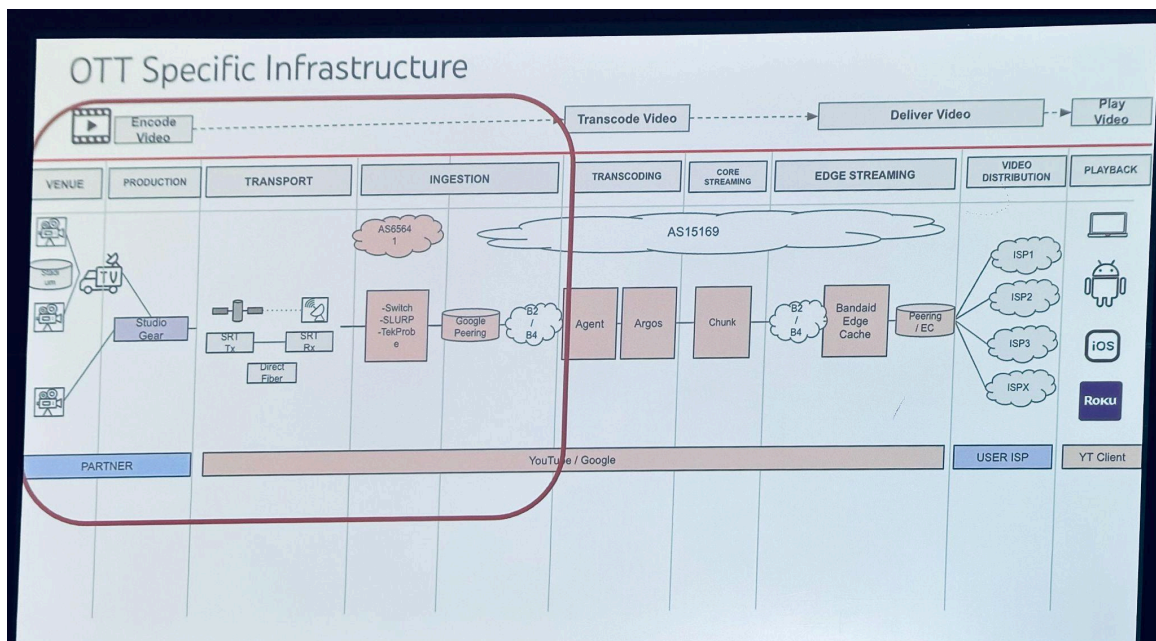


## Team Mission

Deliver the highest quality and most reliable premium live video streams to enable YouTube products.

為支援這些 OTT 產品的營運, YouTube 建構了龐大而複雜的技術體系, 涵蓋影片採集、處理、傳輸、播放等各個環節。

### 3. OTT 基礎設施的演進



圖說: YouTube TV 架構圖。

#### 整體架構

YouTube TV 的基礎設施旨在處理透過網際網路傳遞直播線性電視的複雜性。上面的流程圖, 說明了關鍵的架構組件和設計原則: YouTube TV 系統中的影片, 從初始來源到用戶裝置的完整工作流程。

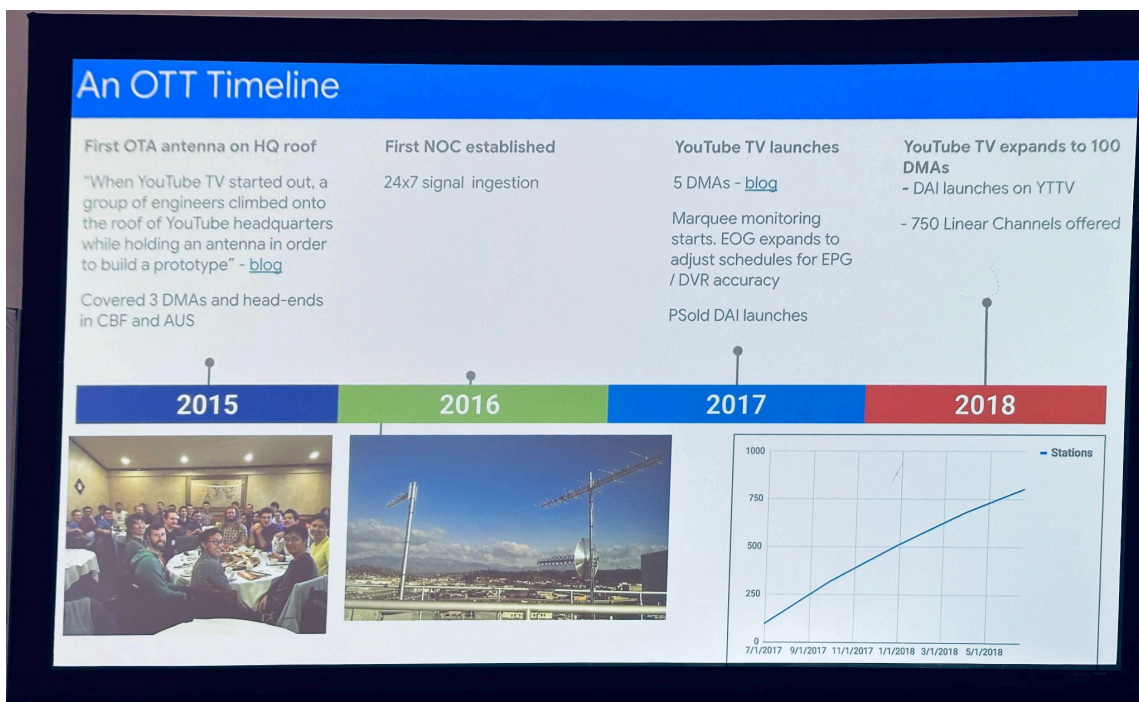
#### 強調區域(紅色框線處)

紅色框線強調了工作流程中的特定部分：從「Encode Video」(影片編碼)，經過「Transport」(傳輸)和「Ingestion」(接收)，到 YouTube/Google 基礎架構內的「Transcode Video」(影片轉碼)。

這突顯了將影片從內容提供者傳輸到 YouTube TV 系統的複雜性和重要性：

- 多樣的輸入格式：處理來自不同合作夥伴的各種輸入格式和協定。
- 可靠的傳輸：確保高品質影片的可靠和低延遲傳輸。
- 網路互連：管理合作夥伴和 YouTube/Google 之間的網路對等互連和交接。
- 大規模接收：同時處理和接收大量的影片串流。

### 3.1 早期階段：



透過網際網路傳遞直播電視，需要一個複雜且強大的基礎設施，以應對即時影片傳輸的獨特需求。在開展 OTT 業務的早期階段，YouTube 面臨著諸多技術挑戰：

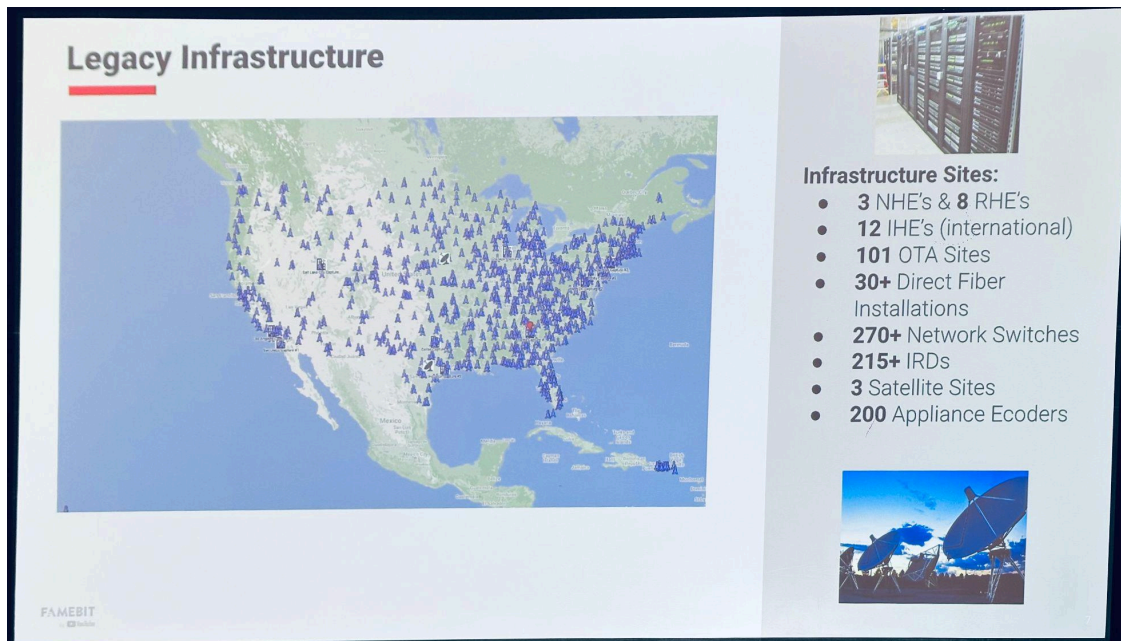
- 廣播級訊號的採集與處理：YouTube TV 與 UGC 內容(也就是我們所理解的社群互動影音平台)所需要的技術是不同的；YouTube TV 是 OTT 平台，需要能夠接收和處理高品質的廣播級訊號。
- 大規模線性直播流的管理：OTT 平台需要支援大量的 24/7 全天候線性直播串流，這對系統的穩定性和可擴展性提出了很高的要求(且與台灣市場所習慣認知的 YouTube 社群影音產品有很大的不同)
- 多樣化的訊號源：YouTube TV 需要從各種不同的訊號源獲取內容，包括 OTA 天線、衛星、光纖等，這增加了系統整合的複雜性。





為應對來自 TV 的挑戰, YouTube 採取了應對 TV 需求的一系列措施, 包括:

- 建設專門的 **OTT** 基礎設施: YouTube TV 建構了一套獨立於 UGC 平台的基礎設施, 以支援 OTT 業務的特殊需求。
- 部署多樣化的訊號採集設備: 為了接收來自不同源的內容, YouTube TV 在各地部署了 OTA 天線、衛星接收設備等。
- 建立光纖連接: 為了獲取高品質的訊號, YouTube TV 與內容提供商建立了大量的光纖連接。

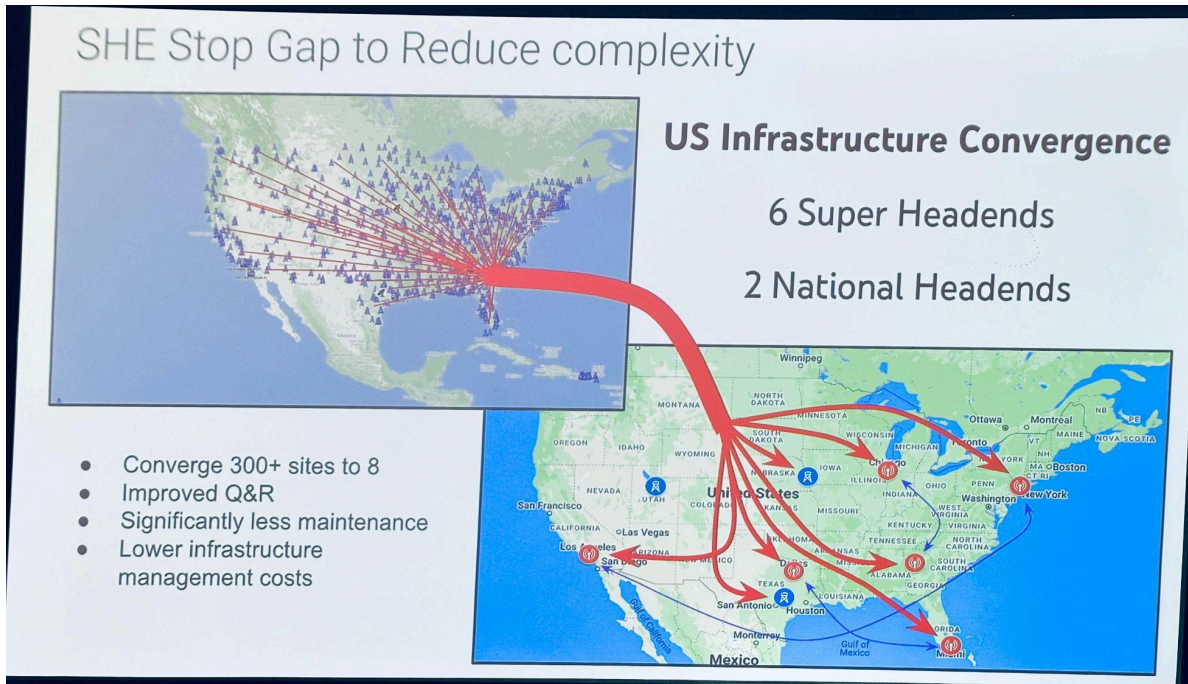




### 3.2 基礎設施的改造：

隨著 OTT 業務的快速發展, YouTube TV 早期的基礎設施已無法支應其所面對的任務, 且逐漸暴露出其技術侷限性：

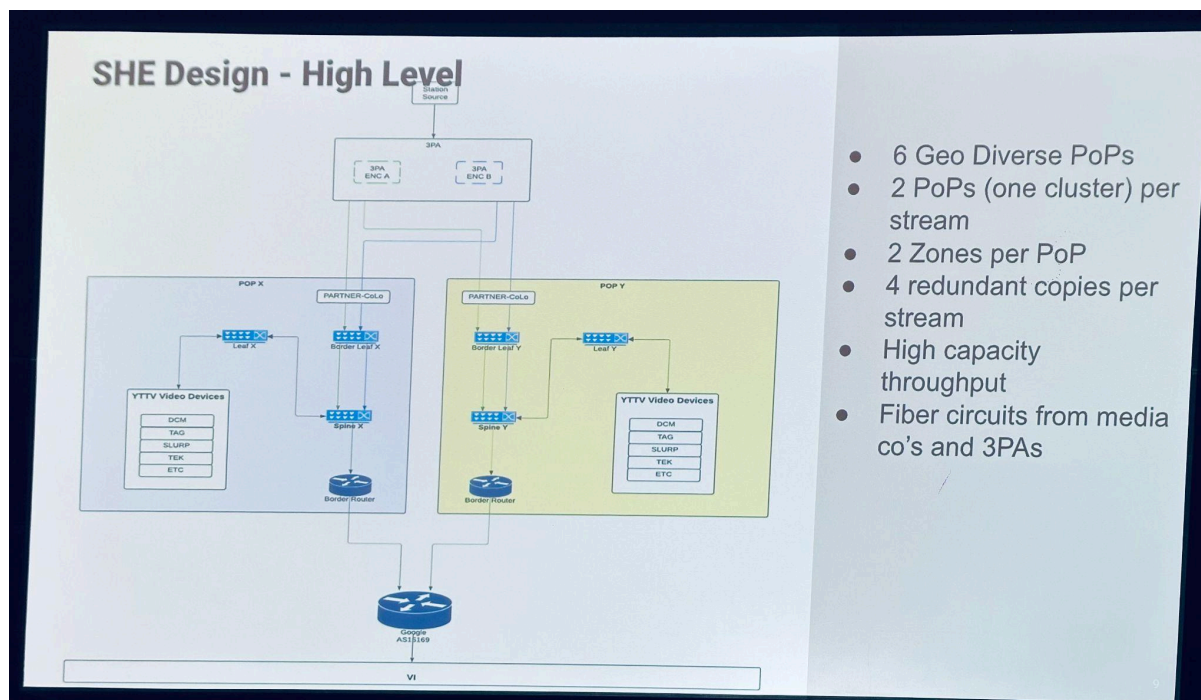
- 技術債：早期建設的基礎設施存在一定的技術債，例如使用了過時的硬體和軟體。
- 維護複雜性：分佈在各地的大量站點，增加了系統的維護複雜性。
- 可擴展性：原有的基礎設施難以滿足 OTT 業務快速增長的需求。



圖說: YouTube TV 全美國佈建超級前端節點(SHE)

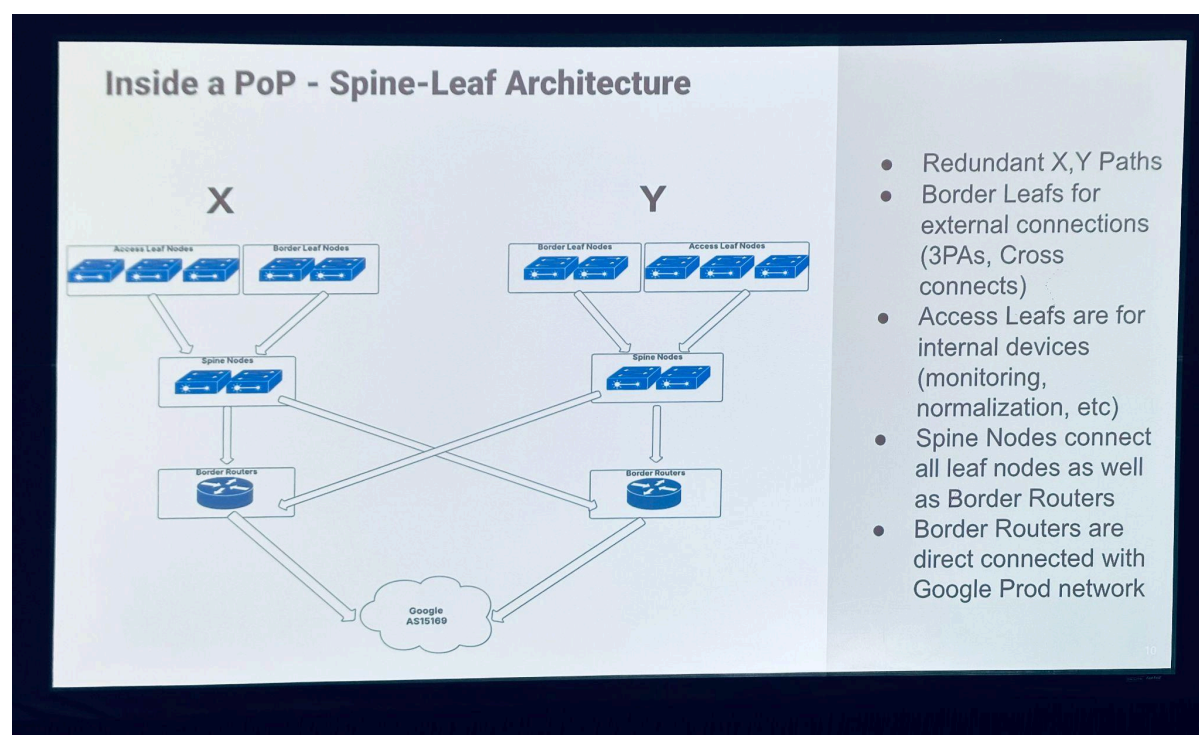
為解決這些問題, YouTube TV 於是啟動大規模的「基礎設施改造」專案, 且多是自有自建模式, 以提高服務技術的自主性：

- 建設超級前端節點：YouTube 將 300 多個分散的站點整合為 6 個超級前端節點(SHE)，這些節點具備更高的處理能力和更大的頻寬。
- 採用光纖連接：新的基礎設施主要依賴光纖連接來傳輸內容，以提高傳輸效率和品質。
- 提高系統的冗餘性：新的基礎設施在設計上充分考慮了冗餘性，以確保系統的高可用性。



圖片說明:「SHE」: Super Head ends

超級前端 (SHE) 是 YouTube TV (以下簡稱 YTTV) 基礎設施的核心元件, 旨在有效且可靠地處理和分配直播電視訊號。它實踐地理位置多樣化。SHE 架構採用分布在不同地理位置的多個 PoP。YouTube TV 和內容合作夥伴之間的連線, 用於接收影片源。每個 PoP 內的「YTTV Video Devices」處理影片處理任務。

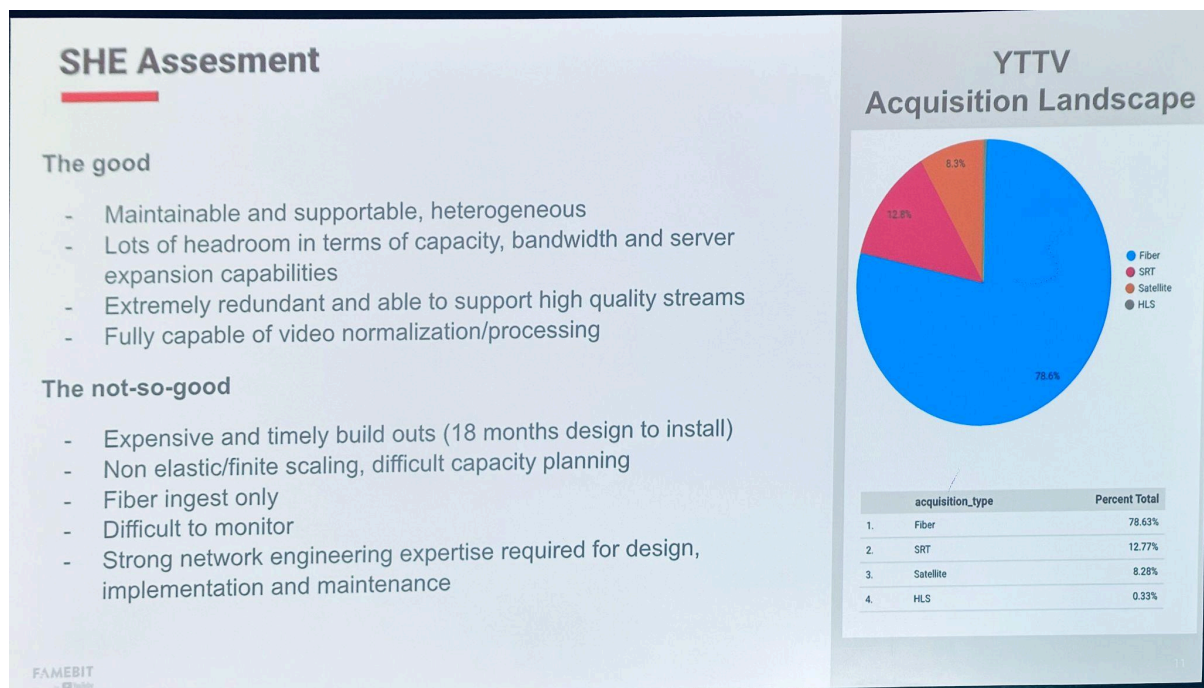


圖說: Spine-Leaf 架構



這是 YouTube TV 使用的「Spine-Leaf 架構」。這是一種現代化的網路設計，通常用於資料中心，以實現高效能和可擴展性。Spine-Leaf 架構的優點在於：

- 高頻寬和低延遲：Spine-Leaf 架構旨在實現高頻寬和低延遲的通訊，這對於處理 PoP 中的大量影片資料至關重要。
- 可擴展性：可以輕鬆新增更多 Leaf 交換器以增加連接的裝置數量，或新增更多 Spine 交換器以增加頻寬容量。這使得 PoP 可以根據需要擴充。
- 冗餘性和容錯能力：冗餘的 X 和 Y 路徑以及架構本身提供了針對故障的恢復能力。
- 可預測的效能：與可能複雜且容易出現瓶頸的舊式網路設計相比，Spine-Leaf 設計通常提供更可預測的效能。



圖說: Spine-Leaf 架構效益評估

總的來說，評估 SHE 架構優點不少：對工程團隊來說，它易於維護和支援，提供充足的容量來處理大量的影片和未來的增長空間；它可以高度冗餘 (redundant)，可確保傳遞高品質的串流，影片處理能力品質優良，可以執行必要的影片處理任務。

雖說如此，卻也不是沒有缺點，像是以下這些缺點：

- 成本和時間：建立 SHE 既昂貴又耗時。
- 可擴展性限制：擴充 SHE 不像基於雲端的解決方案那樣具有彈性。
- 接收限制：它主要依賴於光纖接收，這可能會限制靈活性。
- 監控複雜性：監控系統可能具有挑戰性。
- 專業知識要求：它需要專業的網路工程知識。

## 4. 影片傳輸協定

### 4.1 現有協定：

YouTube 的 OTT 平台支援多種影片傳輸協定，包括：

- **SRT(Secure Reliable Transport)**：一種旨在透過不可靠網路傳輸高品質影片的協定。
- **HLS(HTTP Live Streaming)**：由 Apple 開發的基於 HTTP 的串流媒體傳輸協定。
- **DASH(Dynamic Adaptive Streaming over HTTP)**：一種國際標準化的自適應位元率串流媒體傳輸協定。
- **RTMP(Real-Time Messaging Protocol)**：一種最初由 Macromedia 開發的協定，用於在網際網路上傳輸音訊、影片和資料。

在實際應用中，YouTube 觀察到：

- **SRT** 的應用主要集中在點播直播場景：對於 24/7 全天候的線性直播，SRT 的應用相對較少。
- **RTMP** 仍然被廣泛使用：儘管 RTMP 是一種較老的協定，但由於其在 UGC 領域的普及，YouTube 仍然需要支援該協定。

| Format                        | Pros                                                                                                                                                                                                                                                                                                                                                                              | Cons                                                                                                                                                                                                                         |
|-------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| MPEGTS Multicast              | <ul style="list-style-type: none"><li>• Most widely used in broadcast</li><li>• Typically high throughput and low latency</li><li>• Easy for multi-partner distribution</li></ul>                                                                                                                                                                                                 | <ul style="list-style-type: none"><li>• Physical vulnerabilities (fiber cuts)</li><li>• Not supported in cloud environments (can explore direct connect w/ multicast support)</li><li>• Expensive, fixed bandwidth</li></ul> |
| SRT                           | <ul style="list-style-type: none"><li>• Internet native</li><li>• Resilient (FEC)</li><li>• Open Source</li><li>• Strong ecosystem support/tooling</li><li>• Strong adoption for live events, growing support for linear</li></ul>                                                                                                                                                | <ul style="list-style-type: none"><li>• More difficult server to implement than HTTP</li><li>• Multiple modes/implementations</li><li>• More expensive / difficult to scale</li><li>• No video encryption</li></ul>          |
| CMAF over CDN for Syndication | <ul style="list-style-type: none"><li>• Passthrough opportunity for distributors (no transcoding required)</li><li>• New Codecs already supported (AV1)</li><li>• Full DRM support</li><li>• Extremely cost effective and simple to implement</li><li>• Can converge D2C and partner distribution processing</li><li>• Better geo targeting/content replacement control</li></ul> | <ul style="list-style-type: none"><li>• CMAF packaging not ideal for transcoding</li></ul>                                                                                                                                   |
| Media over Quic (MoQ)         | <ul style="list-style-type: none"><li>• Low latency, target latency</li><li>• Head of line blocking, better congestion control &amp; tput</li><li>• Multi-format support</li><li>• Secure, Scalable</li><li>• Advertise media metadata</li></ul>                                                                                                                                  | <ul style="list-style-type: none"><li>• New emerging standard, low adoption</li></ul>                                                                                                                                        |

圖說：多種影片傳輸協定之優劣點比較

### 4.2 未來方向：

為滿足 OTT 業務不斷發展的需求，YouTube TV 正在積極探索和推廣新的影片傳輸協定：



- **CMAF over CDN**: CMAF (Common Media Application Format) 是一種標準化的容器格式，可以與 CDN (Content Delivery Network) 結合，實現更高效的串流媒體傳輸。(CMAF 也是 TikTok 深入研究的一種格式，後面會提到。)
- **Media over QUIC**: QUIC 是一種由 Google 開發的新的傳輸層網路協定，有望提供更低的延遲和更高的效率。

YouTube 認為，這些新的傳輸協定將：

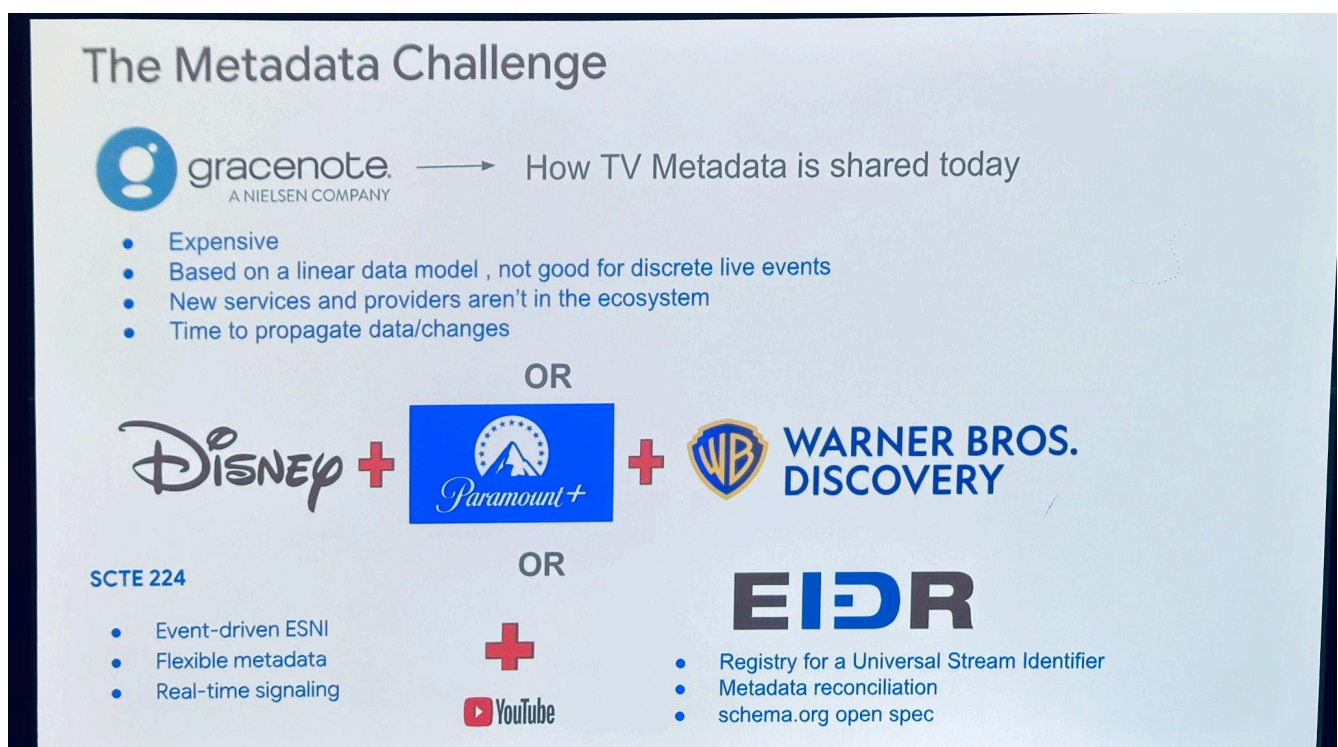
- 降低傳輸成本：新的協定有望取代昂貴的光纖傳輸，從而降低傳輸成本。
- 提高傳輸效率：新的協定能夠更好地利用網路資源，提高傳輸效率。
- 增強靈活性：新的協定能夠提供更大的靈活性，例如更好地支援自適應位元率播放和內容保護。

## 5. Metadata 資料管理

### 5.1 現有方案：

在 OTT 平台中，metadata 資料(例如節目指南資訊)對於使用者體驗至關重要。目前，業界常用的 metadata 資料管理方案包括：

- **Grace Note**：一種提供 metadata 資料服務的供應商。
- 內容提供商的 **API**：一些內容提供商會提供自己的 API，以供 OTT 平台串接並獲取 metadata 資料。



然而，這些方案存在一些問題：

- 成本高昂：使用 Grace Note 等供應商的服務會產生較高的成本。
- 資料模型不靈活：現有的方案可能無法很好地支援 OTT 平台對於 metadata 資料靈活性的需求。
- 資料更新延遲：使用 Grace Note 等供應商的服務，資料更新可能存在延遲。

## 5.2 未來方案：

為了解決現有方案的問題，YouTube 正在探索新的 metadata 資料管理方案：

- **SCTE 224**：一種用於傳輸節目 metadata 資料的標準。
- **Universal Stream Identifier (USID)**：一種用於唯一識別影片串流的標準，本質上是一個分配給媒體串流特定實例的唯一標籤；把它想像成特定影片或音訊串流的指紋。這個識別碼在串流的整個生命週期中保持一致，從它的來源和編碼到它傳遞給最終使用者。透過提供一種標準化且有效的方式來管理和追蹤媒體串流及其相關 metadata 資料，在現代 OTT 媒體傳輸中發揮著關鍵作用。它們的採用對於改善互通性、自動化工作流程以及確保複雜 OTT 生態系統中的資料準確性至關重要。它會在以下幾個工作項目產生影響，如節目指南 (EPG)：確保準確顯示節目資訊；權利管理：執行內容權利和限制；廣告：傳遞目標廣告並追蹤廣告效果；以及資料分析：收集有關收視率和 QoE (體驗品質) 的資料。

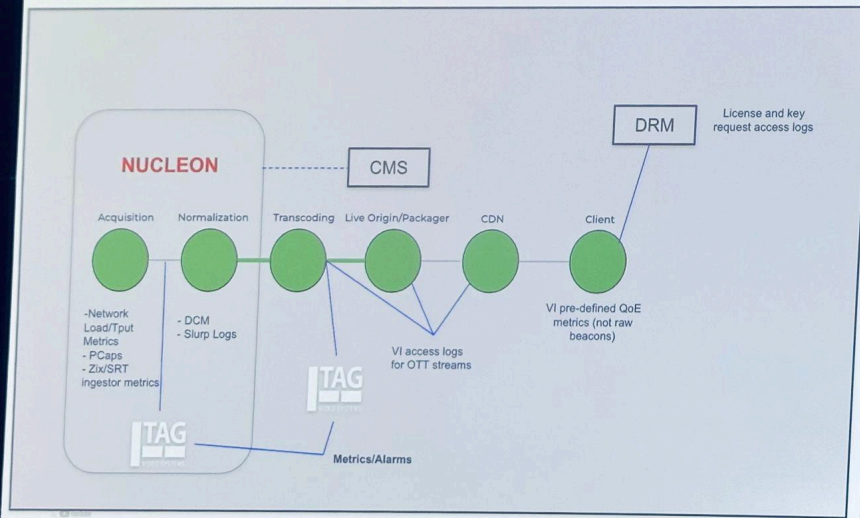
YouTube 認為，結合使用這些標準，建構一個開放的 metadata 資料管理方案，從而：

- 降低成本：採用開放標準可以降低對單一供應商的依賴，從而降低成本。
- 提高效率：標準化的 metadata 資料格式可以提高資料交換的效率。
- 增強互操作性：採用開放標準可以增強不同系統之間的互操作性。



## Observability Challenge

### A Need for End-to-End Monitoring



### Challenges

- Correlating ingestion events with QoE playback issues
- Monitoring the bitflow of a channel/stream – distributed request tracing
- Differentiating component, transport level and content level issues
- Realtime Raw log collection

### Goals

- Realtime – high volume, high dimensionality, high cardinality, real-time data pipelines (sub 30 seconds)
- Fast querying
- Curated, well understood metrics for operations
- End to end observability

## 6. 可觀測性

### 6.1 挑戰：

在 OTT 影片傳輸過程中，實現點對點的串流健康狀態可視化非常重要。然而，傳統的廣播訊號採集方式，給串流健康狀態的監控帶來了諸多挑戰：

- 多樣化的採集方式：由於訊號採集方式的多樣化，很難對整個傳輸鏈路進行統一的監控。
- 缺乏串流識別 (**Stream ID**)：缺乏串流識別，使得難以追蹤影片串流在整個傳輸過程中的狀態。
- 資料關聯性：採集到的資料可能難以與用戶端的播放品質指標（例如緩衝、播放失敗）建立關聯。這一點，事關優化客戶服務品質提升提升會員數量。

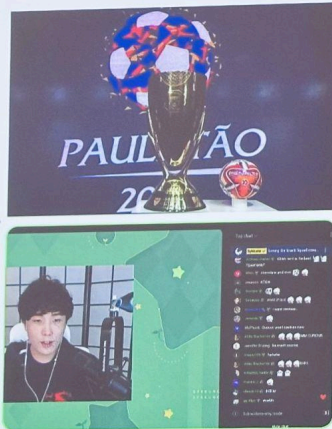
### 6.2 解決方案：

為了解決這些挑戰，YouTube 正在探索以下解決方案：

- 建立資料池：透過建立資料池，可以整合來自不同來源的監控資料，實現對整個傳輸鏈路的統一監控。
- 採用統一的串流識別 (**Stream ID**) 機制：透過串流識別機制，可以追蹤影片流在整個傳輸過程中的狀態。
- 擴展監控指標：YouTube 正在探索擴展監控指標，以更好地反映影片品質，例如 VMAF 和 PSNR。

## Diversity and Scale

### YouTube Live: from casual mobile to professional 4K HDR



Streams range from Mobile to Gaming and Desktop to Broadcast.

Beyond YouTube TV, sports are a big international presence on YouTube.

YouTube Live Streams can be running for years

## 7. 影片處理流程: YouTube 創作者生態系統中的技術考量

來自YouTube工程團隊資深成員的講者 Kirk Caller 也提到了 YouTube 如何處理來自創作者的直播內容，以及平台在技術層面上面臨的挑戰和解決方案。其中，創作者扮演著核心角色，他們的內容多樣性以及技術使用習慣，都深刻地影響著 YouTube 的技術策略。

- RTMP 協議與創作者的遷移

Kirk Caller 指出，RTMP 協議作為一種較舊的、基於 Flash 的格式，雖然已經存在許久，但仍然是許多創作者使用的主要直播協議。他強調，要讓創作者放棄他們已經習慣的 RTMP 設定並轉向新的協議，是一項極具挑戰性的任務（例如 YouTuber 們必須配合平台改變或擴充自身的硬體設備，就會增加創作者的營運成本而可能促使他們轉台，改去不同的平台創作內容，導致創作者流失），必須要有充分的理由驅使他們改變。因此，YouTube 在技術發展的過程中，必須同時考量到對現有 RTMP 協議的支援，以及對新技術的推廣，雙管齊下。

- Enhanced RTMP 與新技術的支援

為了在支援現有創作者的同時，也能夠推動新技術的發展，YouTube 選擇與 Adobe 和 Amazon 合作，共同開發了 Enhanced RTMP 協議。這一新的協議在保留了 RTMP 基本架構的同時，也加入了對 AV1、HEVC、HDR 等新一代編碼格式和技術的支援。這使得創作者能夠在不改變原有工作流程的前提下，也能夠利用更先進的技術來提升直播內容的品質。

- 創作者頻寬限制的考量

Kirk Caller 也提到了創作者在進行直播時，可能會受到頻寬限制的問題。許多創作者可能在家中或工作室進行直播，這些環境的網路條件可能遠不如專業的製播室，更不要說集合廣播



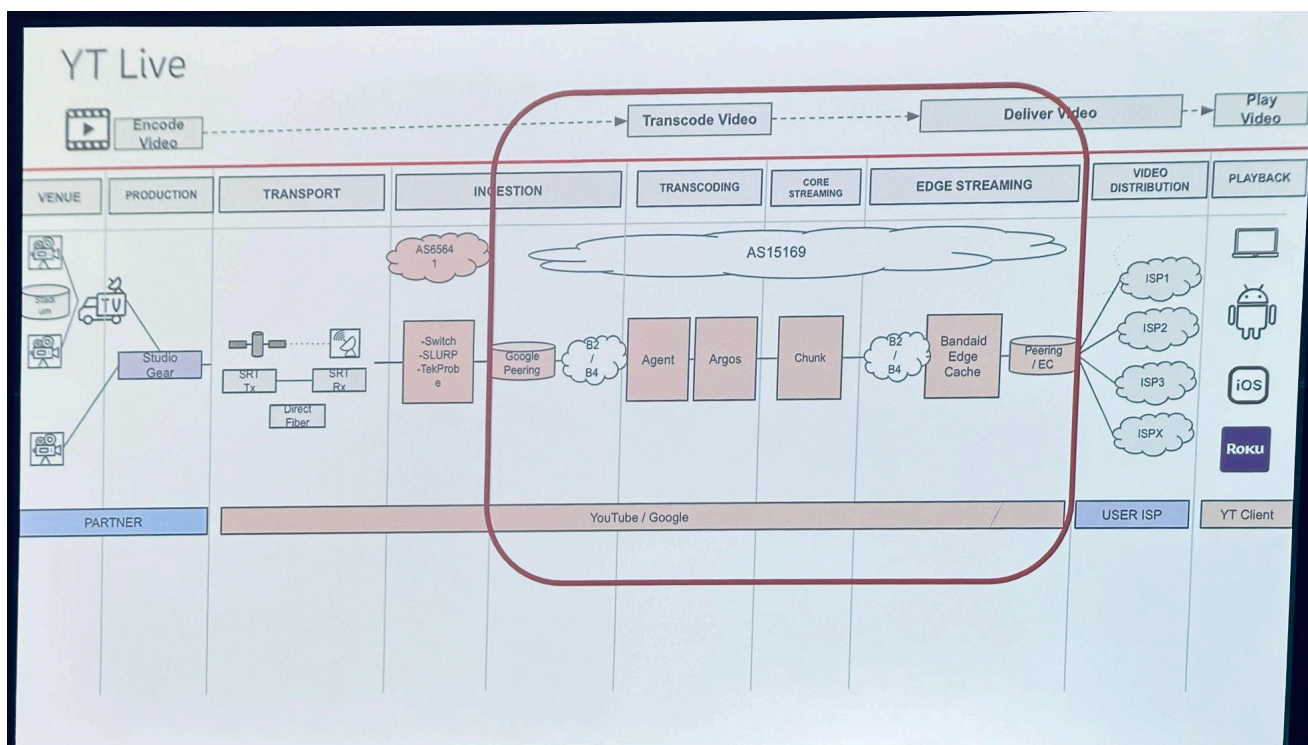
級專業之大成的電視台了。因此，YouTube 的技術設計必須能夠做到，容納各種不同網路條件下，都能做到內容傳輸，並提供匹配不同環境下的相應解決方案，例如自適應碼率串流等，以確保觀眾能夠獲得流暢的觀看體驗。

- YouTube 平台內容的多樣性與規模

Kirk Caller 強調了 YouTube 平台內容的多樣性與規模，這也是平台在技術上面臨的一大挑戰。YouTube 上的直播內容，不僅數量龐大，而且形式多樣，從持續數年的長期直播，到只有幾分鐘的短暫直播都有。內容的類型也包羅萬象，既有手機拍攝的垂直影片，也有專業製作的 4K HDR 內容。這種內容的多樣性，要求 YouTube 的技術架構必須具備高度的靈活性和可擴展性，才能夠有效地處理來自不同創作者的各種內容。

- 創作者市場的潛力

Kirk Caller 認為，創作者是一個具有巨大潛力的市場。隨著直播技術的普及和發展，越來越多的創作者開始嘗試製作高品質的直播內容，這也帶動了對更先進的影音技術和設備的需求。因此，對於那些希望進入 OTT 市場的供應商來說，創作者是一個不容忽視的重要客群。也就是說，若從技術的角度而言，從技術力創作力高的「創作者端」，到「技術力高專業度高」的廣播級訊號，YouTube TV 都必須能夠一體兼顧。這不但是一個極大的技術考驗，更是市場考驗。



圖說: YouTube LIVE 架構

## 7.1 影片採集：

綜合前述，YouTube 的 OTT 平台必須能支援多種影片採集格式，包括 SRT、HLS、DASH 和 RTMP。其中，RTMP 仍然是 UGC 內容的主要採集格式。

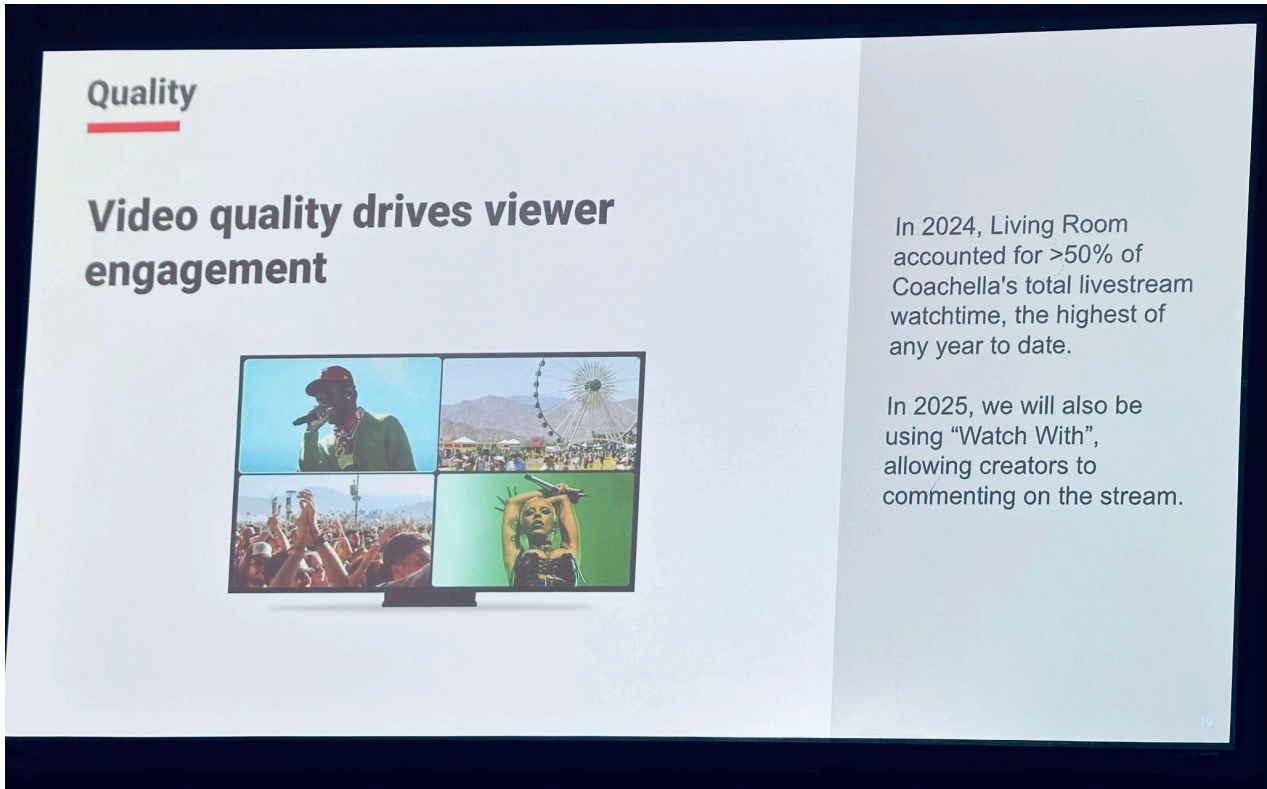


為了更好地支援 RTMP, YouTube 與 Adobe 和 Amazon 合作, 推出了 Enhanced RTMP 協定, 該協定支援 AV1、HEVC、HDR 等新的影片編碼格式。

## 7.2 影片轉碼:

YouTube 會對採集的影片進行轉碼, 以適應不同的播放裝置和網路條件。轉碼的目標是:

- 提高播放品質: 透過轉碼, 可以提高影片的清晰度和流暢度。
- 降低頻寬消耗: 透過轉碼, 可以產生不同位元率的影片流, 以適應不同的網路條件。



**Quality**

**Video quality drives viewer engagement**

In 2024, Living Room accounted for >50% of Coachella's total livestream watchtime, the highest of any year to date.

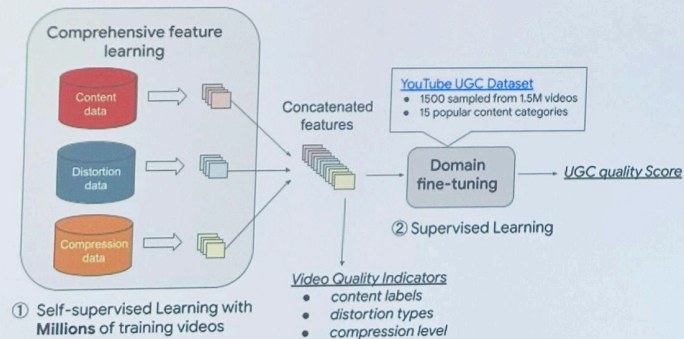
In 2025, we will also be using "Watch With", allowing creators to commenting on the stream.

19

## 7.3 影片品質評估:

為了客觀地評估影片品質維持穩定, YouTube 開發了一種名為 UVQ (YouTube User Video Quality) 的監督式機器學習模型。

## Measuring Audio/Video Quality



In 2022, YouTube and Google research opened source UVQ

- ML Model trained on Mean Opinion Scores (MOS)
- Based on content, distortions and compression data
- “No reference” metric

圖片說明：測量音訊/影片品質：UVQ

在 2022 年，YouTube 和 Google 研究部門開放原始碼 UVQ。這是一項測量音訊/影片品質的方法，用於評估使用者產生內容 (UGC) 影片的品質。

- **ML 模型以平均意見分數 (MOS) 進行訓練：**機器學習 (Machine Learning, ML) 模型使用平均意見分數 (Mean Opinion Scores, MOS) 進行訓練。MOS 是媒體感知品質的數值度量，通常是從人類觀眾的主觀評估中獲得的。
- **基於內容、失真和壓縮資料：**模型使用這三種類型的資料來預測品質。
  - 內容資料：指的是原始的影片和音訊資料本身。(例如，流派、類別)
  - 失真資料：有關影片或音訊中的瑕疵或損傷的資訊(例如，模糊、雜訊)。
  - 壓縮資料：這包括有關使用的壓縮技術及其參數的資訊(例如，位元速率、編解碼器)。
- **「無參考」指標：**這是至關重要的特徵。這意味著它可以在不需要原始影片的情況下評估品質，使其適合於評估 UGC。在不將處理後的影片與原始、原始版本進行比較的情況下，也可以評估品質。這對於 UGC 場景至關重要，因為原始影片通常不可用。

UVQ 模型可以：

- **預測用戶滿意度：**UVQ 模型可以預測用戶對於影片品質的滿意度。
- **支援無參考評估：**UVQ 模型是一種無參考評估方法，這意味著它不需要原始影片作為參考。
- **考慮內容特性：**UVQ 模型會考慮影片內容的特性，例如運動量，從而更準確地評估影片品質。

這項技術對於 YouTube 不斷優化觀眾體驗並且有助於留客(retention)來說非常重要，因為它可以大規模監控上傳影片的品質，識別並可能對品質不佳的影片採取行動，優化影片處理和傳遞。

| UVQ score          | Perceptual quality               |
|--------------------|----------------------------------|
| [1.0, 3.5)         | relatively low                   |
| [3.5, 4.1)         | fair                             |
| [4.1, 5.0]         | relatively high                  |
| 0.05~0.1 UVQ delta | Just noticeable difference (JND) |

| MOS score on TV | Satisfaction | Resolution |
|-----------------|--------------|------------|
| 4.17            | 95%          | 1080       |
| 4.0             | 95%          | 720        |

圖片說明: 如何測量音訊/影片品質

這張圖延續了關於「品質」的討論，並以 UVQ 分數的概念為基礎。它還將這些指標與感知品質和觀看者滿意度聯繫起來。

舉例：電視上的 **MOS** 分數、滿意度和解析度：

- 底部的表格將平均意見分數 (MOS)、觀看者滿意度和影片解析度 (特別是電視觀看) 聯繫起來。
- 電視上的 **MOS** 分數：
  - 4.17 的 MOS 分數與 1080 解析度的 95% 滿意度相關。
  - 4.0 的 MOS 分數與 720 解析度的 95% 滿意度相關。
- 滿意度：這表示將影片品質評為令人滿意的觀看者百分比。
- 解析度：這是指影片的垂直解析度 (1080 像素或 720 像素)。

這個表格將客觀品質指標 (MOS)、主觀觀看者評分 (滿意度) 和電視上的影片解析度聯繫起來的經驗資料。就畫面解析度來舉例說明：即使 MOS 分數略有不同，觀看者通常對電視上的 1080 和 720 解析度都非常滿意 (95%)。所以也就是說，能夠藉此指標得出，目前市場上用戶最大宗的滿意規格，藉以權衡使用者的品質需求與開發團隊技術量能，以達到最佳邊際效益。



這些資料可以為以下方面的決策提供資訊：

- 目標解析度：為不同的裝置和觀看情境選擇合適的解析度。
- 品質權衡：在頻寬限制下平衡品質。

此資訊對於 YouTube TV 來說至關重要，可以：

- 建立品質基準和閾值。
- 最佳化影片處理和傳遞，以最大程度地提高觀看者的滿意度。
- 針對解析度和品質權衡做出明智的決策。

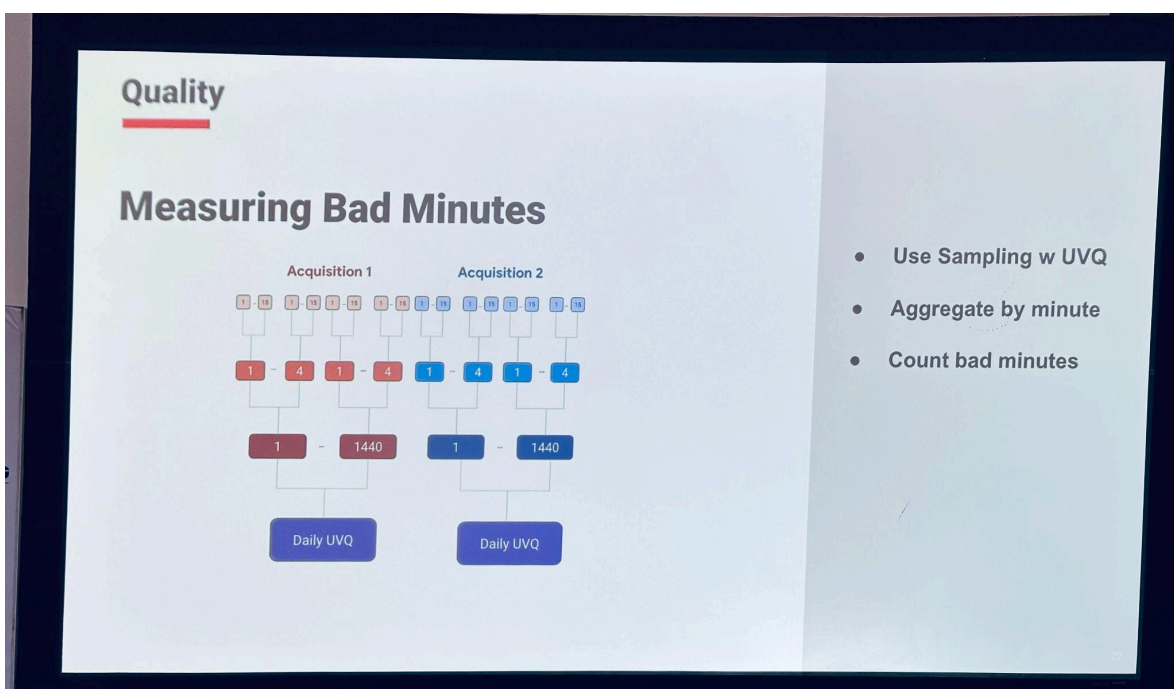
綜上所述，我們就可以理解，即便現在4K拍攝已經滿地都是，但是真正的4K內容數量，成長卻相對有限。反映在YouTube平台上，亦是如此。

## 8. 用戶體驗指標

### 8.1 關鍵指標：

「用戶體驗」，是衡量 OTT 平台成功與否的關鍵指標。YouTube 關注的關鍵用戶體驗指標包括有：

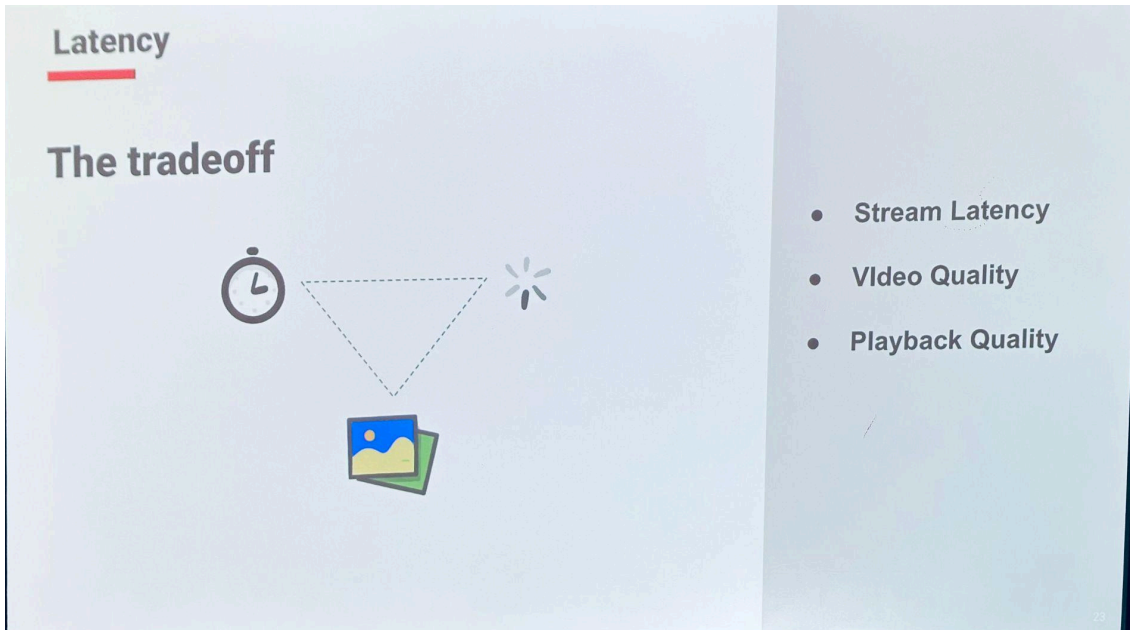
- 啟動延遲：從用戶發起播放到影片開始播放的時間。
- 播放失敗率：影片播放失敗的比例。
- 緩衝率：影片播放過程中發生緩衝的頻率。
- 位元率：影片播放的位元率，通常作為影片品質的代理指標。



## 8.2 評估方法：

為了更全面且深入地評估用戶體驗，而且在多數情況下，又不能取得原始影片或是重現當下播放環境條件以作對照，因此YouTube 採用以下方法：

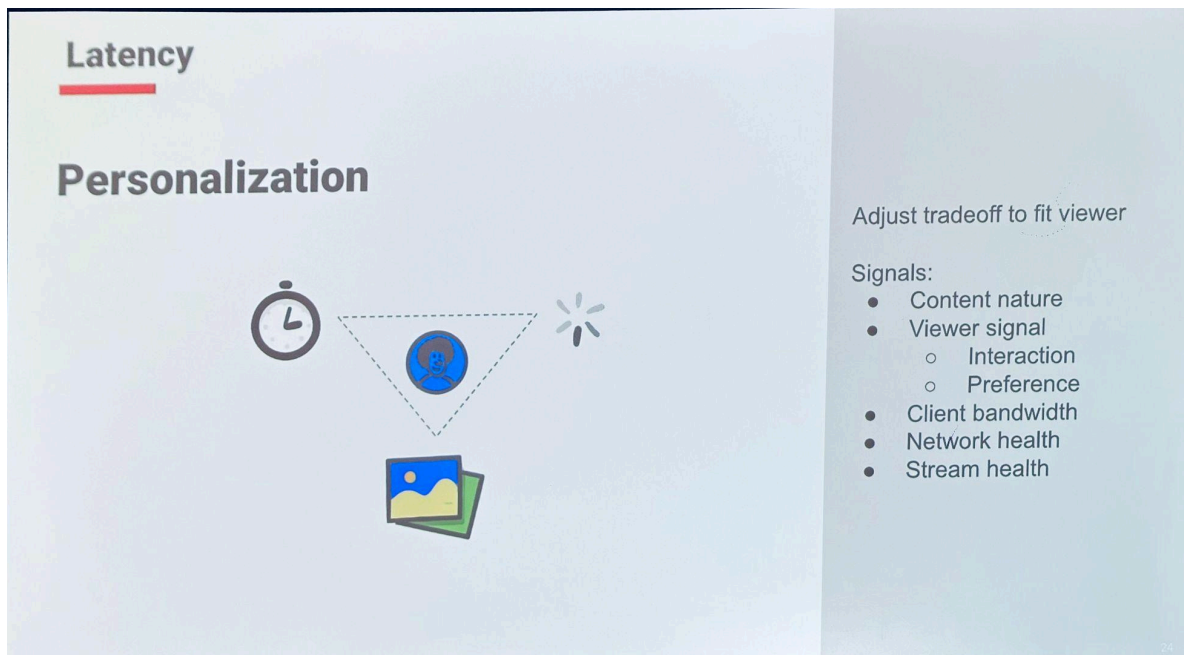
- 合成體驗指標：YouTube 使用合成體驗指標，例如「啟動延遲超過 1 秒的播放比例」，來衡量用戶體驗。
- 生存分析：YouTube 使用生存分析方法，來評估用戶在遇到問題（例如緩衝）後繼續觀看影片的可能性。



## 8.3 優化策略：

為了持續優化用戶體驗，YouTube 採取以下策略：

- 設定明確的改進目標：YouTube 會設定明確的用戶體驗指標改進目標，並將其納入 OKR (Objectives and Key Results) 體系。
- 監控和分析用戶體驗資料：YouTube 會持續監控和分析用戶體驗資料，以識別潛在的問題和改進機會。
- 優化影片播放流程：YouTube 會不斷優化影片播放流程，例如減少啟動延遲、降低緩衝率，以提升用戶體驗。



## 9. OTT 平台中的 QoE 分析與評判方法

在 OTT (Over-The-Top) 流媒體服務領域, QoE (Quality of Experience, 用戶體驗質量) 是衡量服務成功與否的至關重要指標。它不僅關乎影片的技術質量, 更涉及用戶在觀看過程中的整體感受和滿意度。YouTube 作為全球領先的影片平台, 高度重視 QoE, 並發展出一套完善的方法來評估和優化用戶的觀看體驗。

### 9.1 QoE 的概念與重要性

- QoE 的定義

QoE 是一個綜合性的概念, 它反映了用戶在使用服務時的整體滿意程度。在 OTT 影片服務中, QoE 不僅僅取決於影片的客觀質量參數(例如解析度、位元率), 還受到多種主觀因素的影響, 包括:

- 影片內容: 內容是否吸引人、是否符合用戶的興趣。
- 服務易用性: 平台的操作是否簡便、導航是否清晰。
- 個人期望: 用戶對於服務的期望值。
- 情境因素: 用戶的觀看環境(例如裝置、網路狀況)。

- QoE 的重要性

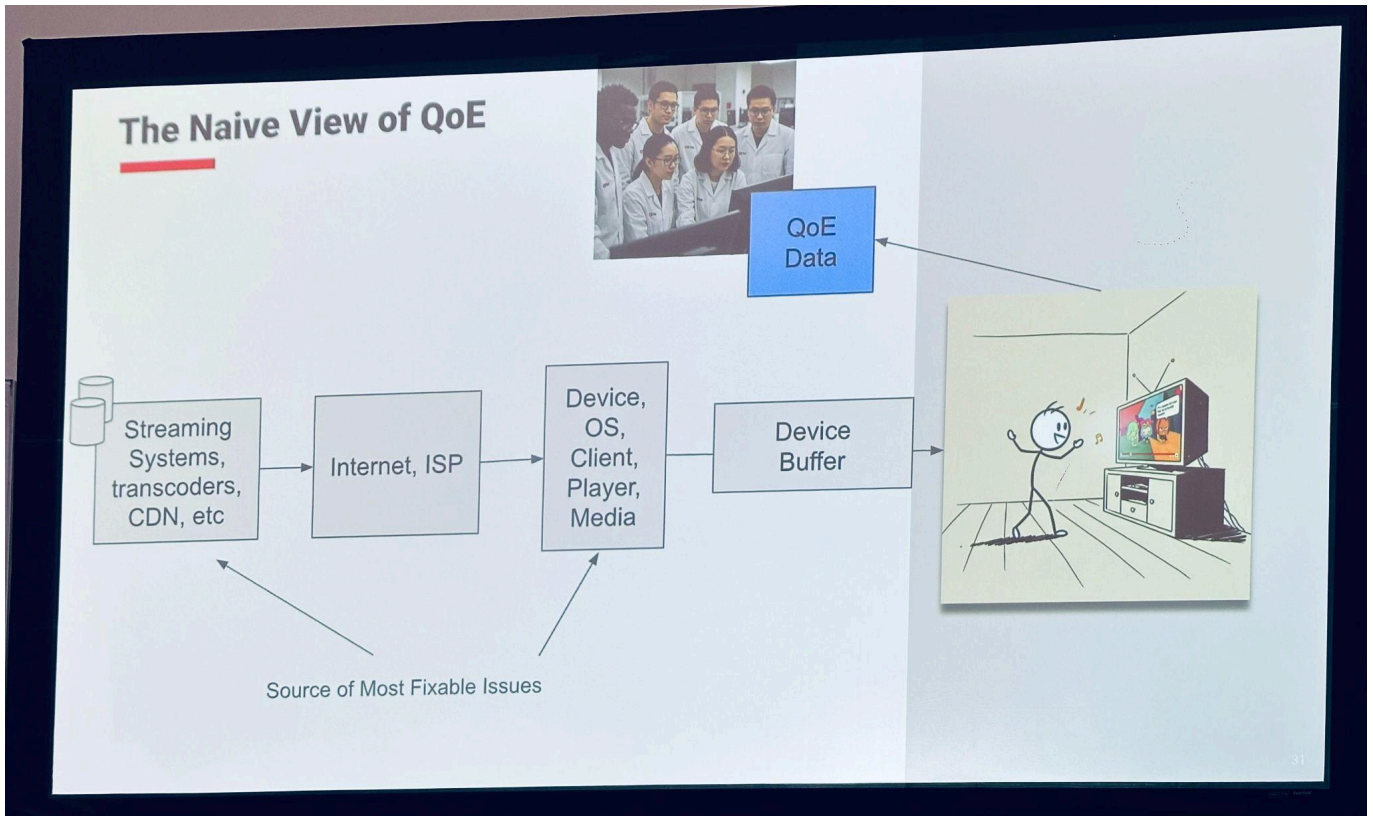
在競爭激烈的 OTT 市場中, 提供優良的 QoE 至關重要。高 QoE 可以帶來諸多好處, 包括:

- 提高用戶黏著度: 滿意的用戶更有可能繼續使用該服務。
- 減少用戶流失: 低劣的 QoE 會導致用戶轉向其他平台。
- 提升品牌形象: 優質的 QoE 有助於建立良好的品牌聲譽。
- 增加商業收益: 高 QoE 可以促進用戶付費意願和廣告點擊率。



## 9.2. YouTube 中 QoE 的評判方法

來自 YouTube 工程團隊的講者們，多次提及 QoE，並詳細介紹了 YouTube 用於評估和改進 QoE 的多種方法。講者 Chaz Masten 在演講中提到了一種 "naive view" 的 QoE 評估方法，這種方法側重於一些傳統的技術指標。



這些指標包括：

- 影片啟動失敗率：影片無法成功開始播放的比例。
- 播放失敗率：影片在播放過程中發生錯誤而停止的比例。
- 緩衝率：影片播放過程中發生緩衝的頻率。
- 位元率：影片播放的位元率，通常作為影片品質的代理指標。
- 啟動延遲：從用戶點擊播放到影片開始播放的時間。

這些指標反映了影片播放的流暢性和穩定性，是 QoE 的重要組成部分。然而，講者也指出，僅僅關注這些指標是不夠的，因為它們無法全面反映用戶的主觀感受。

## 9.3 合成體驗指標

為了更全面地評估 QoE，YouTube 採用了一種稱為 "synthetic experience metrics" (合成體驗指標) 的方法。這種方法的核心思想是將不同的 QoE 指標進行時間歸一化，以便於比較和分析。

## Synthetic Experience Metrics

- A Synthetic Experience Metric is a metric built out of QoE metrics time-normalized in some way to allow comparison and improvements. It can include combining multiple categories of QoE, or showing the likelihood of a bad experience.
- Simple: % of video starts > 1000ms 7DA
- Complex: Survival-based
- Hire a Data Scientist to build these

合成體驗指標的例子包括：

- 啟動延遲超過閾值的播放比例：例如，計算啟動延遲超過 1 秒的播放佔總播放量的比例。
- 一段時間內的平均緩衝次數：例如，計算每分鐘或每小時的平均緩衝次數。

透過使用合成體驗指標，YouTube 可以更方便地比較不同時間段、不同內容、不同用戶群體之間的 QoE，從而更好地發現問題和評估改進效果。

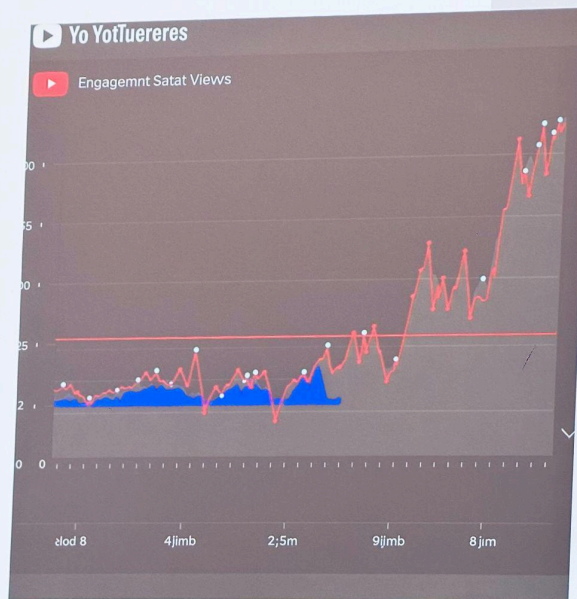
### 9.4 什麼是 SLO？

SLO 代表 **Service Level Objective** (服務等級目標)。

**SLO 的目的與功能**

- SLO 是用於定義和衡量服務可靠性和效能的重要工具。
- 它們為特定的指標設定了目標等級，表明可接受或期望的效能。
- 如果服務的效能低於定義的 SLO，則表示存在需要關注的問題。
- SLO 有助於團隊確定工作重點，並根據數據做出有關服務改進的決策。

## Synthetic Experience Metrics - What is an SLO?



### 9.5 如何使用 SLO

- SLO 可以應用於各種指標，例如延遲、緩衝率或任何其他使用者體驗或系統健康狀況的指標。
- 透過追蹤相對於 SLO 的效能，團隊可以監控服務隨著時間推移的可靠性。
- SLO 也可用於定義「錯誤預算」，它表示與目標效能的可接受偏差量。
- 超出錯誤預算可能會觸發警報或需要採取特定措施來解決根本問題。

SLO 提供了一種本質清晰且可衡量的方式來定義、追蹤和維持服務所需的品質和可靠性。



## Synthetic Experience Metrics

**Metric SLOs *either* protect user experience, or allow you to set improvement goals.**

- Rebuffer is an excellent metric to set **defensive** SLOs on, but it's difficult to set improve goals on due to changes in audience and networking conditions.
- You end up having to spend a lot of time comparing things by hand, not a great use of engineering time.
  - MTBR slightly better
  - Likelihood to rebuffer within x minutes the best...

34

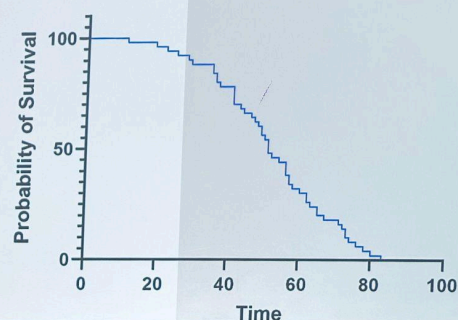
### 9.6 生存分析

講者 Chaz Masten 介紹了一種更高級的 QoE 評估方法, 即 "survival analysis" (生存分析)。生存分析最初是醫學領域用於研究病人存活率的方法, 後來被應用於多個領域, 包括 QoE 評估。

## Synthetic Experience Metrics p2

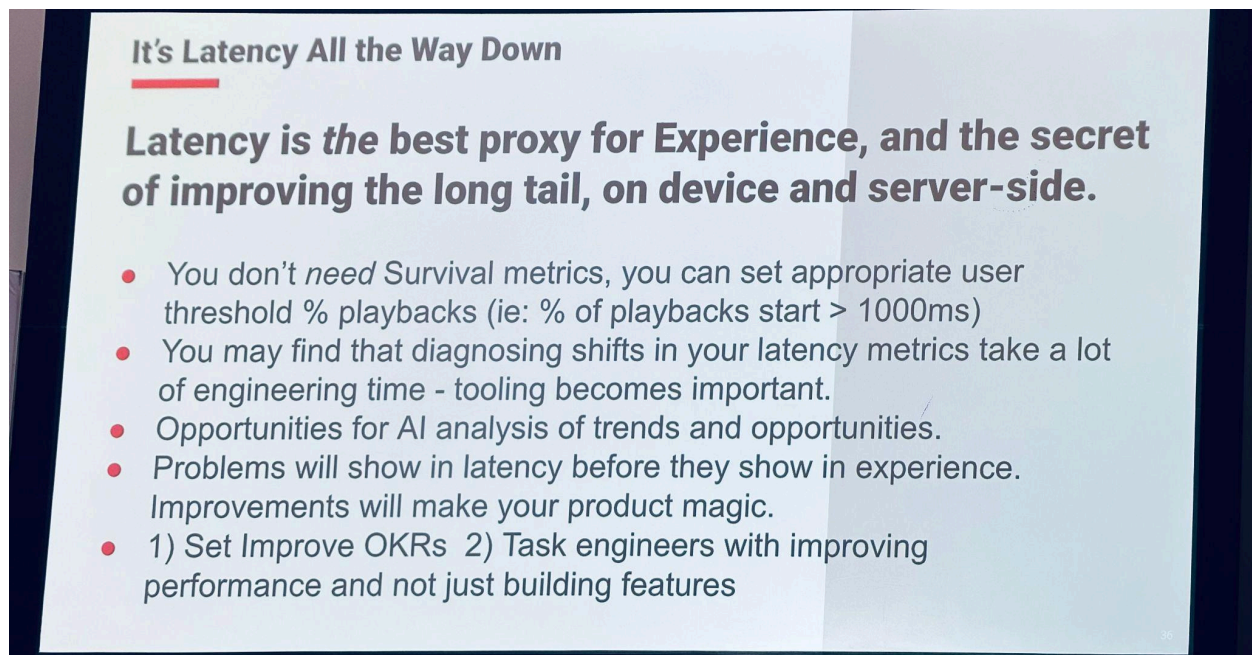
**Metrics based on Survival Curves let you make intelligent tradeoffs with other Defensive Metrics.**

- Survival metrics measure the likelihood of surviving a Bad User Experience event over time (latent startup event, rebuffer, fatal error)
- Create a human perceptible threshold
  - >1s is "slow"
  - % of playbacks per day going over that threshold
- Sessions with a Slow Playback Rate



35

在 OTT 影片服務中，生存分析可以幫助評估用戶在遇到負面體驗（例如啟動延遲過高、發生緩衝）後繼續觀看的可能性。生存分析的結果通常以 "生存曲線" 的形式呈現，該曲線顯示了用戶在不同時間點的 "存活率"（即繼續觀看的比例）。



**It's Latency All the Way Down**

**Latency is the best proxy for Experience, and the secret of improving the long tail, on device and server-side.**

- You don't *need* Survival metrics, you can set appropriate user threshold % playbacks (ie: % of playbacks start > 1000ms)
- You may find that diagnosing shifts in your latency metrics take a lot of engineering time - tooling becomes important.
- Opportunities for AI analysis of trends and opportunities.
- Problems will show in latency before they show in experience. Improvements will make your product magic.
- 1) Set Improve OKRs 2) Task engineers with improving performance and not just building features

生存分析的優勢在於，它可以更全面地反映用戶的行為模式和容忍度，從而更準確地評估 QoE。例如，如果生存分析顯示，大量用戶在啟動延遲超過 5 秒後就選擇離開，那麼平台就需要優先優化啟動延遲。

## 9.7 UVQ 模型

前面(段落7.3)有提到，為了更客觀地評估影片質量，YouTube 運用"UVQ"(YouTube Video Quality)的監督式機器學習模型：

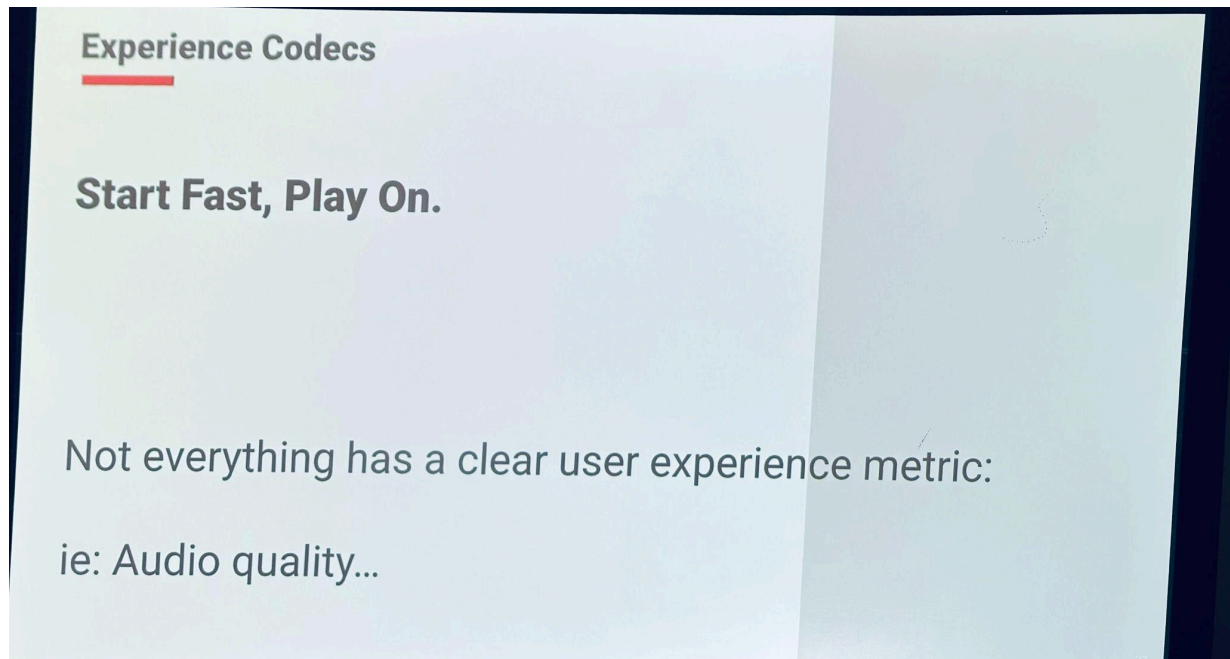
- 預測用戶對影片質量的感知：UVQ 模型會綜合考慮影片的內容失真和壓縮數據，預測用戶對影片質量的滿意度。
- 支援無參考評估：UVQ 模型是一種 "no reference metric"，這意味著它不需要原始影片作為參考，可以直接對接收到的影片進行質量評估。
- 考慮內容特性：UVQ 模型會根據影片內容的特性（例如運動量）選擇不同的模型，以提高評估的準確性。

UVQ 模型為 YouTube 提供了客觀、準確的影片質量評估手段，有助於優化影片編碼和傳輸流程，提升用戶的觀看體驗。因此，UVQ也是 QoE的重要工具之一。

## 9.8 其他 QoE 因素

除了上述方法，講者們還提到了其他一些影響 QoE 的重要因素，包括：

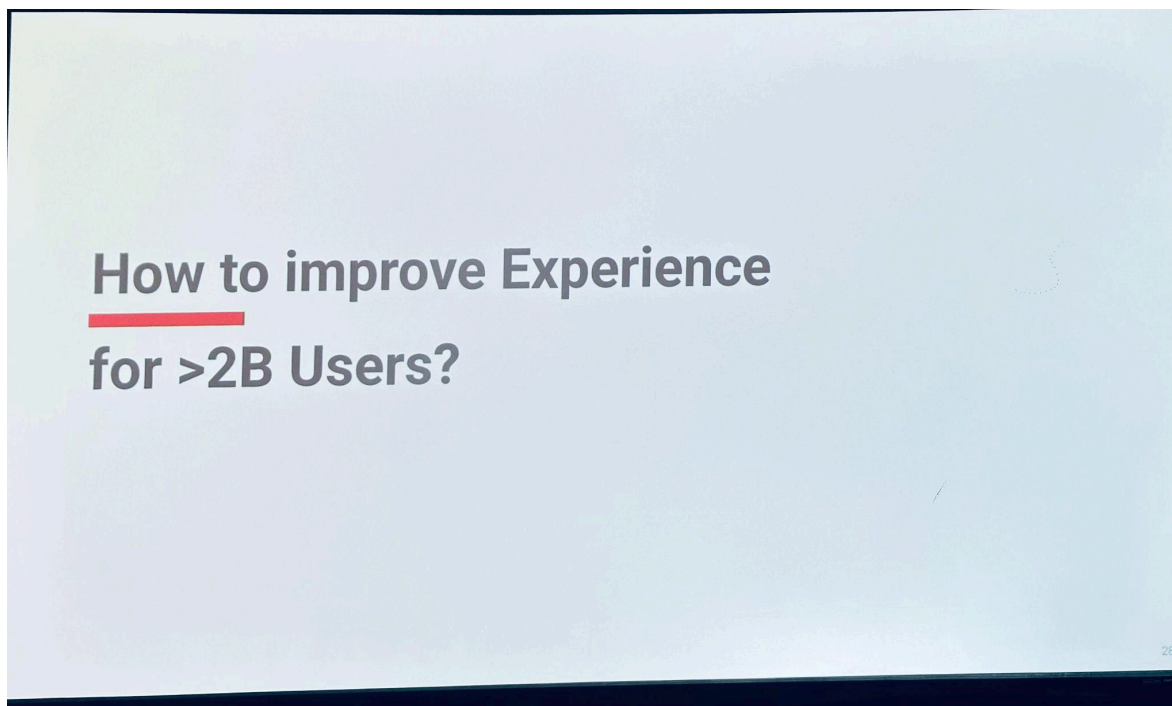
- 延遲：對於直播內容，低延遲至關重要。講者 Kirk Caller 強調了延遲在不同應用場景下的不同重要性，例如，在視訊會議中，延遲是最重要的因素，而在體育直播中，則需要在延遲和影片質量之間做出權衡。
- 音訊品質：講者 Stephen Robertson 指出，音訊的穩定性和響度對於用戶體驗有著重要影響。YouTube 開發了 "stable volume" (穩定音量) 功能，以解決音量突變和廣告音量過大的問題，從而提升音訊方面的 QoE。



QoE 是 OTT 平台成功的關鍵。為了準確評估和持續優化 QoE, YouTube 採用了多種方法, 包括傳統的 QoE 指標、合成體驗指標、生存分析、UVQ 模型, 以及對延遲和音訊品質的關注。這些方法相互補充, 共同構建了一個全面、客觀的 QoE 評估體系。透過深入理解和有效管理 QoE, OTT 平台可以提升用戶滿意度, 增強用戶黏著度, 並在競爭激烈的市場中取得成功。



## 10. 音訊品質管理



### 10.1 響度問題：

音訊響度是影響用戶體驗的另一個重要因素。OTT 平台常常面臨以下響度問題：

- 音量突變：不同內容之間的音量可能存在差異，導致用戶需要頻繁調節音量。
- 廣告音量過大：廣告的音量通常高於正片，這會給用戶帶來不適。

其中「廣告片」是一個重要因素。實際的情況是，為了吸引觀眾的注意力，廣告商往往無所不用其極，甚至會冒險犧牲觀看體驗，故意把音量響度推大，目的是吸引在廣告時段忽略的觀眾的耳朵，讓他們多停留注意力，即使一兩秒也好。

### 10.2 解決方案：

為了解決這些問題，YouTube 採取了以下措施：

- 響度標準化：YouTube 會對音訊進行響度標準化，以確保不同內容之間的音量保持一致。
- 穩定的音量功能：YouTube 推出了穩定的音量功能，以自動調節音量，減少音量突變帶來的影響。
- 用戶控制：YouTube 允許用戶關閉穩定的音量功能，以滿足用戶的個人化需求。

**Table D.1 Reference Sound Pressure Level**

| Categories                                 | Room Volume in Cubic Feet                                                                                                                                                                        | SPL in dB re 20 $\mu\text{N/m}^2$ |
|--------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|
| I, II<br>(Mix rooms)                       | > 20,000                                                                                                                                                                                         | 85*                               |
|                                            | 10,000 < 19,999                                                                                                                                                                                  | 82                                |
|                                            | 5,000 < 9,999                                                                                                                                                                                    | 80                                |
|                                            | 1,500 < 4,999                                                                                                                                                                                    | 78                                |
|                                            | < 1,499                                                                                                                                                                                          | 76                                |
| III (Edit rooms sometimes used for mixing) | Depends on room usage. For editing purposes, may be controlled by the editor for use with the material at hand. For final program mixing, follow the recommendations for categories I, II above. |                                   |
| IV (Booths, vans)                          | < 1,500                                                                                                                                                                                          | 76                                |
| V (Headphones)                             | Use 2 cc. Coupler and set 400 Hz level to 78 dB.                                                                                                                                                 |                                   |
| * Per SMPTE RP 200 [6]                     |                                                                                                                                                                                                  |                                   |

### 10.3 持續改進：

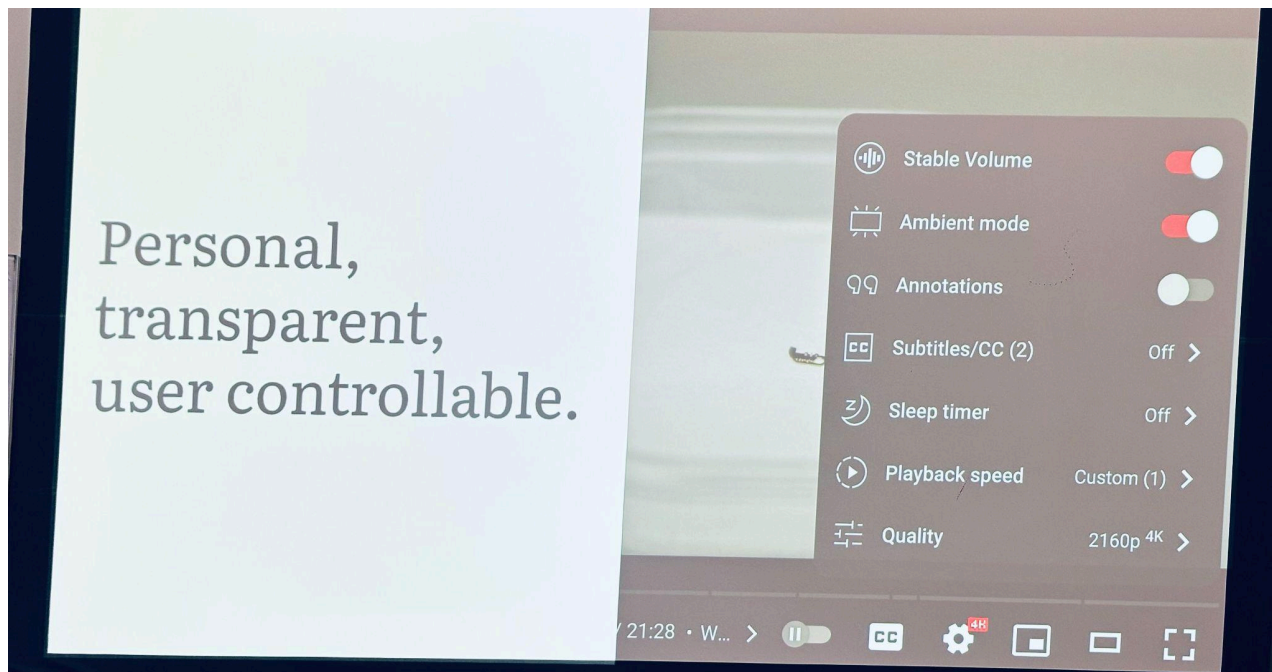
YouTube 致力於持續改進音訊品質管理：

- 更智慧的演算法：YouTube 正在開發更智慧的演算法，以更精確地控制音訊響度。
- 更多用戶選項：YouTube 計劃為用戶提供更多音訊相關的選項，例如對話增強。

Client-controlled DRC is not new  
(especially in Living Room).

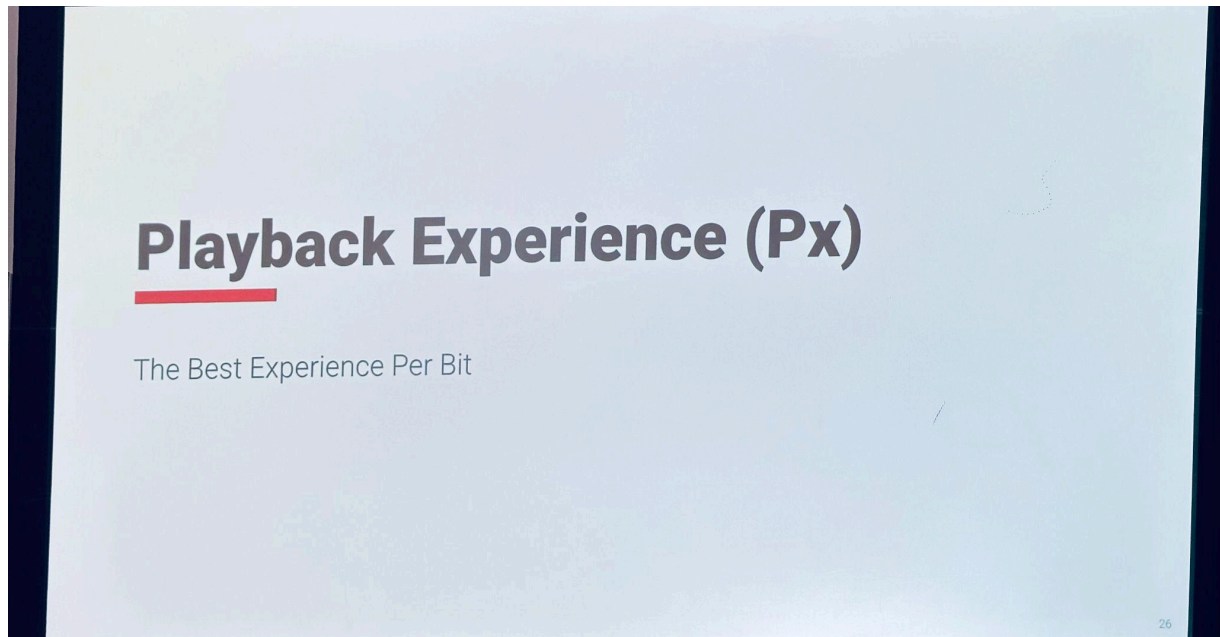
## 11. 小結

YouTube 在建構和營運大規模 OTT 服務方面，累積了豐富的技術經驗。透過不斷的技術創新和優化，YouTube 致力於為用戶提供高品質、可靠的 OTT 影片串流服務。本報告詳細介紹了 YouTube 在 OTT 基礎設施、影片傳輸協定、元資料管理、可觀測性、影片處理流程、用戶體驗指標和音訊品質管理等方面的技術實踐，希望能為 OTT 產業的技術發展提供有益的借鑒。





## 第二章 OTT 平台中廣告音量與效益的平衡策略



### 1. 引言

在 OTT (Over-The-Top) 平台中，廣告是重要的營收來源。然而，廣告音量的大小直接影響用戶體驗。過大的廣告音量會引起用戶反感，降低用戶體驗質量 (QoE)，甚至對廣告效益產生負面影響。因此，如何在廣告效益和用戶體驗之間取得平衡，找到最佳的廣告音量，是 OTT 平台運營的重要課題。本報告將深入分析 YouTube 在廣告音量管理方面的策略，以及評判最佳廣告音量的方法。

### 2. 廣告音量對用戶體驗的影響

#### 2.1 過大音量的負面影響

YouTube 工程團隊講者 Stephen Robertson 明確指出，過大的廣告音量會對用戶體驗產生顯著的負面影響。具體來說，過大的廣告音量會導致：

- 用戶不適：突然的音量增大會刺激用戶的聽覺，引起不適感，降低觀看體驗。
- 內容干擾：廣告音量過大會破壞節目內容的連貫性，影響用戶對內容的沉浸感。
- 用戶反感：長期遭受過大廣告音量的轟炸，會使得用戶對該平台和廣告內容、甚至廣告主產生反感情緒。

## Loudness variation was a top user complaint

### 2.2 對於用戶行為的影響

更重要的是，過大的廣告音量不僅僅停留在感官層面的不適，還會直接影響用戶在平台的行為。Stephen Robertson 提到，YouTube 的數據清晰地顯示，過大的廣告音量會導致：

- 廣告跳過：用戶會更傾向於跳過音量過大的廣告，影響所及，降低了廣告的曝光度和有效性。
- 內容放棄：如果廣告無法跳過，用戶可能會直接放棄觀看當前內容，甚至離開平台，造成用戶流失。
- 工作階段放棄：在更嚴重的情況下，用戶可能會直接放棄當前工作階段，減少在平台上的停留時間和互動。

## 3. 廣告音量對廣告效益的影響

### 3.1 廣告跳過

如前所述，過大的廣告音量會增加用戶跳過廣告的可能性。這對廣告效益產生直接的負面影響，因為：

- 曝光度降低：廣告被跳過意味著用戶沒有完整地觀看到廣告內容，從而降低了廣告的曝光度。
- 訊息傳達失敗：廣告的主要目的是向用戶傳達產品或服務的訊息。如果廣告被跳過，則訊息傳達失敗，廣告目標無法實現。

### 3.2 品牌形象受損

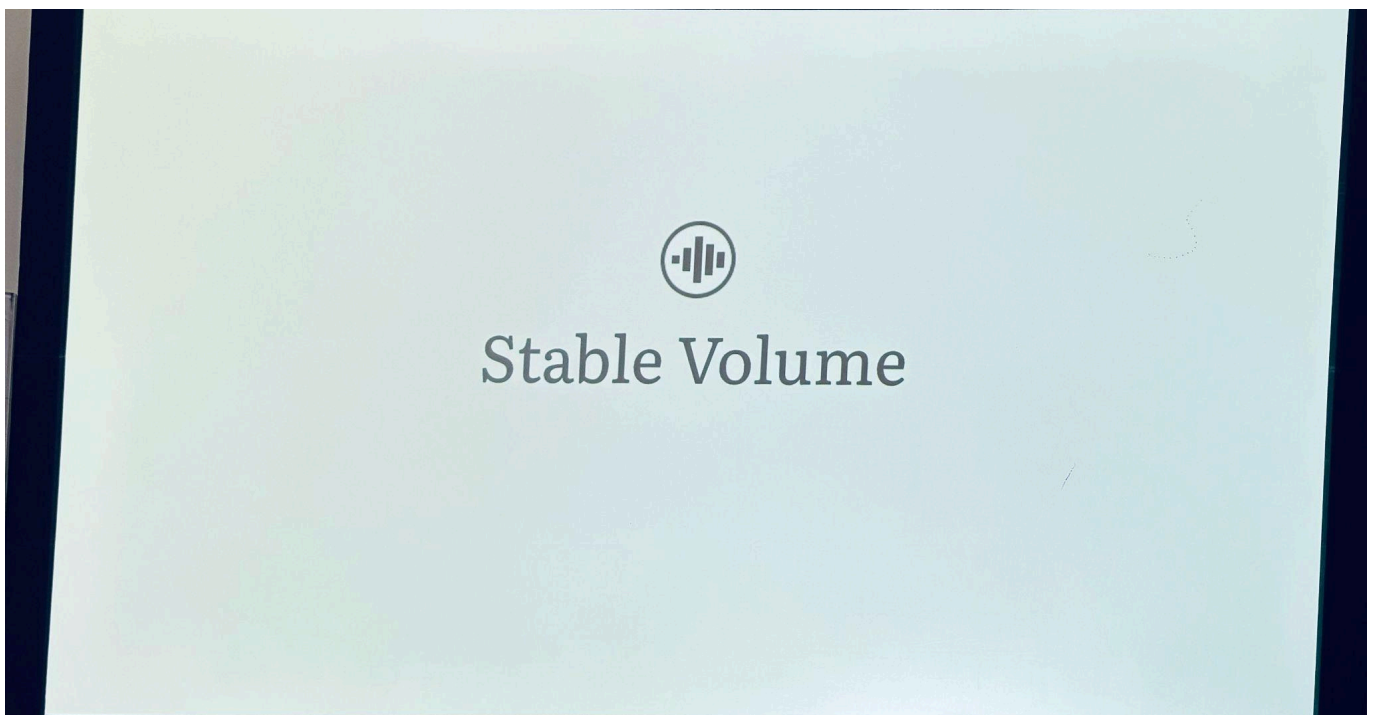
除了直接的曝光度降低，過大的廣告音量還會對品牌形象造成損害。這是因為：

- 負面聯想：過大的音量會讓用戶將廣告與負面體驗聯繫起來，例如煩躁、不適。這種負面聯想會轉移到廣告所宣傳的品牌上，損害品牌形象。
- 信任度降低：用戶可能會認為使用過大音量來吸引注意力的品牌，缺乏對用戶體驗的尊重，從而降低對該品牌的信任度。

### 3.3 廣告效益降低

綜合以上因素，過大的廣告音量最終會導致廣告效益的降低。這是因為：

- 點擊率下降：由於用戶對廣告的反感和跳過行為，廣告的點擊率可能會下降。
- 轉換率降低：即使部分用戶觀看了廣告，負面的品牌形象也可能降低他們的購買意願，從而降低廣告的轉換率。



## 4. 評判最佳廣告音量的方法

### 4.1 用戶回饋

Stephen Robertson 強調了用戶回饋在優化音量控制策略中的重要性。具體來說，可以透過以下方式收集用戶回饋：

- 直接投訴：用戶可以透過平台的投訴管道，直接反映對廣告音量的意見。
- 使用行為分析：分析用戶在觀看廣告時的行為，例如跳過率、離開率等，從而間接了解用戶對廣告音量的反應。
- 問卷調查：定期進行用戶問卷調查，收集用戶對廣告音量的偏好和容忍度。



## 4.2 指標監控

除了用戶回饋，還可以透過監控一系列指標來評估廣告音量對用戶體驗和廣告效益的影響。這些指標包括：

- 廣告跳過率：監控用戶跳過廣告的比例，以評估廣告音量是否導致用戶反感。
- 內容放棄率：監控用戶在廣告播放期間放棄觀看內容的比例，以評估廣告音量是否影響用戶留存。
- 用戶停留時間：監控用戶在平台上的平均停留時間，以評估廣告音量是否影響用戶的整體使用行為。
- 廣告點擊率：監控廣告的點擊率，以評估廣告音量是否影響用戶的互動意願。
- 廣告轉換率：監控廣告的轉換率，以評估廣告音量是否最終影響廣告的商業效益。

## 4.3 A/B 測試

A/B 測試是一種常用的優化策略，也可以用於評判最佳廣告音量。具體來說，可以將用戶隨機分為不同的組別，每組用戶觀看不同音量的廣告，然後比較各組用戶的行為和反饋，從而找出最佳的廣告音量。

## 4.4 演算法優化

為了更精確地控制廣告音量，YouTube 正在開發更智慧的演算法。這些演算法可以：

- 自動調節音量：根據不同的內容和用戶情境，自動調節廣告音量，以達到最佳的平衡。
- 考慮心理聲學：考慮人類對聲音的感知特性 (心理聲學)，例如等響曲線，從而更準確地控制用戶對音量的感知。
- 個人化音量設定：允許用戶設定自己的廣告音量偏好，以提供更個人化的體驗。

## 5. 小結

在 OTT 平台中，廣告音量的大小是影響用戶體驗和廣告效益的重要因素。過大的廣告音量會引起用戶反感，降低廣告效益。為了找到最佳的廣告音量，OTT 平台需要綜合考慮用戶回饋、指標監控、A/B 測試和演算法優化等多種方法。透過精細化的音量控制策略，OTT 平台可以在廣告營收和用戶體驗之間取得平衡，實現可持續發展。

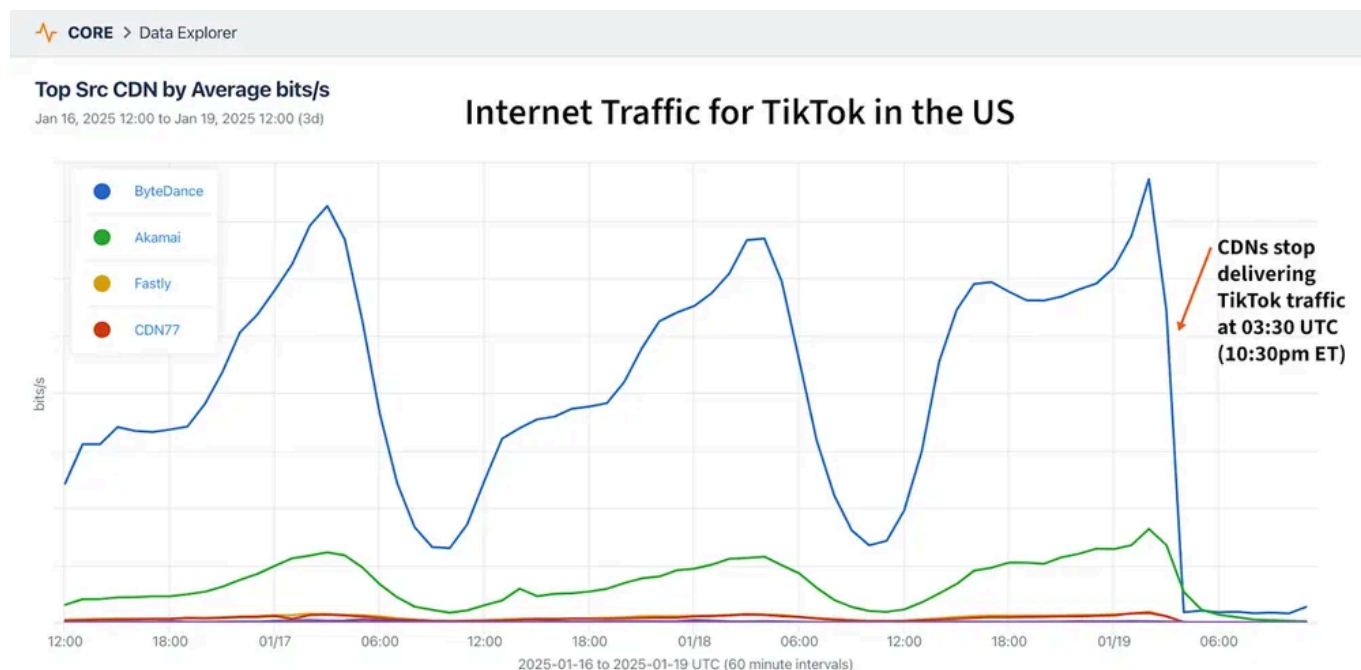
## 第三章 TikTok使用CMAF的低延遲經驗

前情提要：

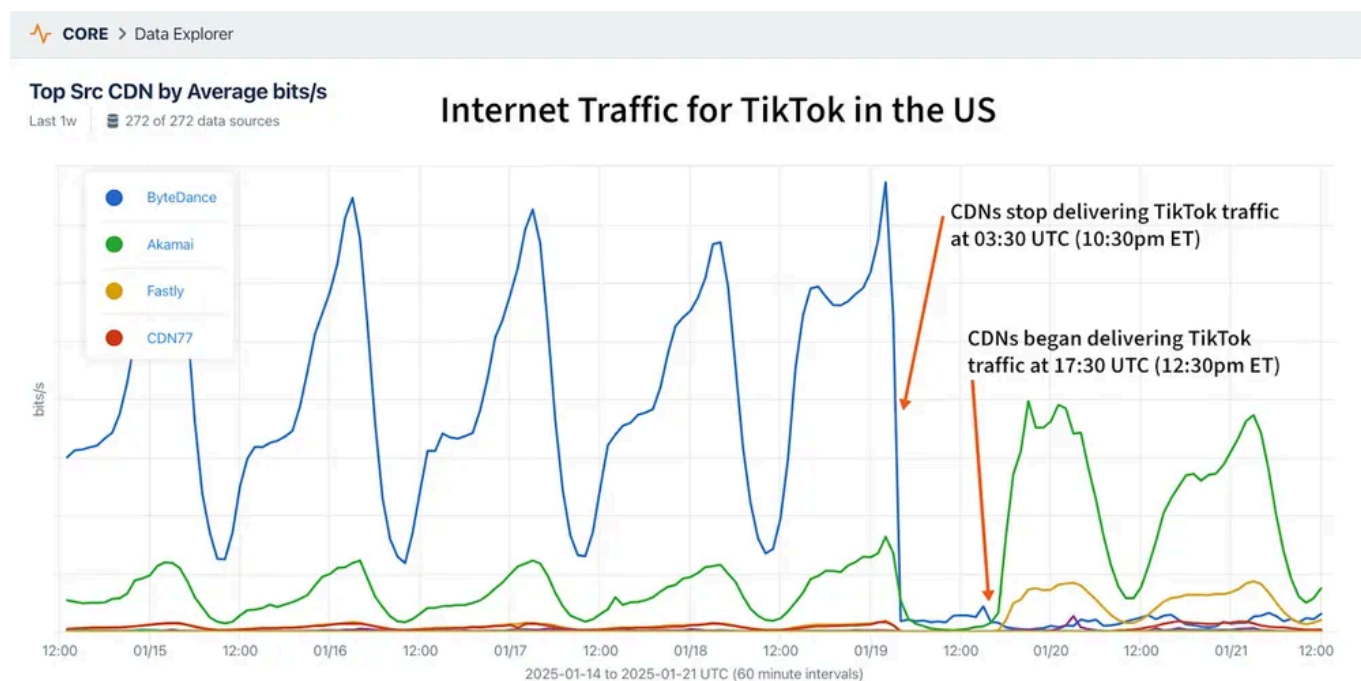
TikTok 在美國市場擁有約 1.7 億美國用戶。2024 年 1 月 19 日，美國最高法院最終裁定 TikTok 的禁令將生效。這導致 TikTok 在當天晚些時候中斷服務，形同在美國「消失」達 14 小時。

根據 Kentik 的 OTT 服務流量追蹤數據顯示：TikTok 在 1 月 19 日凌晨 03:30 停止流量傳

輸。此時，原本TikTok所使用的母公司字節跳動的 CDN，流量完全消失。這一變化清晰地體現在 Kentik 數據平台上，流量的驟降如同「斷崖」。



流量直到19 個小時之後才恢復。此時可以看到，流量結構也發生了顯著變化：過去由母公司字節跳動 CDN 提供的大部分流量，現已被 Akamai、Fastly 和 CDN77 等第三方 CDN 所取代。這不僅是一項技術調整，更是 TikTok 剝離其母公司資源的重要步驟。



數據也顯示，TikTok 的流量切換並未影響其用戶使用體驗。流量切換至第三方 CDN，意味著 TikTok 在技術架構上，已經進一步本地化，同時也可能帶來更嚴格的數據隱私和合規，以符合美國政府的要求。這可能也是為未來獨立運營或等待被美國本土資金收購創造條件。

直播串流是TikTok的一項重要業務。他們所使用的第三方CDN之中，也包括了一家名為「CDN77」供應商所提供的服務。雙方在低延遲直播技術上有深度合作。他們將 5 秒以下的任何內容標記為「低延遲」，將 1 秒以下的任何內容標記為「超低延遲」。若將這些數字對標到以下情境更能看出差異：透過 HTTP 進行的常規直播延遲通常為 20 到 40 秒，而數位電視廣播大約需要 4 到 6 秒才能到達觀眾。在一些案例中，低延遲表現至關重要。例如社群媒體的主播需要能夠在聊天中及時對觀眾做出反應。如果發生 30 秒的延遲，那麼保持串流媒體的吸引力是不可能的。大型體育賽事也是如此。沒有人希望在聽到鄰居的歡呼聲之後才親眼看到進球。CDN77 透過 CMAF 格式的優化技術，協助 TikTok 提高流量效能並節省成本。



**STREAMING SUMMIT AT NABSHOW** THE OTT EVENT OF THE YEAR #NABSTREAMINGSUMMIT

**Case Study, TikTok and CDN77: How to Optimize CMAF for Low-Latency Live Streaming**

Learn about optimizations made to the TikTok live streaming platform, including those applied at the origin, CDN edge, and adjustments made to the player to manage CPU load.

**Tomas Bacik**  
VP, Client Solutions, CDN77

**YingAn Wang**  
Engineering Leader, Live Streaming Infrastructure, TikTok

**CDN77** **TikTok**

## 1. 引言：TikTok 的直播串流平台

雖然 TikTok 以其短影音點播內容而廣為人知，但它也提供強大的直播串流功能。

- **低延遲重點：**TikTok 直播串流的一個關鍵特點是強調低延遲，目標是實現 4 秒以下的點到點延遲。在直播購物等特定使用場景中，TikTok 致力於實現更低的延遲（低於 3 秒），以增強互動性和使用者體驗。
- **內容多樣性：**TikTok 的直播串流平台包含多樣化的內容，包括直播購物、教育、新聞、遊戲、美容和時尚。



- 使用者互動：使用者主要透過他們的「動態消息」與直播互動，該消息整合了短影音和直播內容。這種基於動態消息的互動模式影響了直播啟動時間最小化的重要性，因為使用者傾向於在直播未能快速開始時滑動切換。
- 多人連線和共同主持功能：的平台使內容創作者能夠使用多人連線和共同主持功能，邀請其他使用者一起直播。

## 2. 低延遲的挑戰與 CMAF 的角色

直播過程中總是需要權衡利弊得失，特別是在以下方面取得平衡：

- 穩定性：確保可靠且不間斷的串流。
- 延遲：盡可能縮短直播事件與呈現給觀眾之間的延遲。
- 成本：管理大規模傳輸直播的相關費用。

TikTok 對低延遲的重視，受到直播購物等使用案例以及維持使用者參與度的需求所驅動，帶來了巨大的技術挑戰。

- FLV 的沿用：由於 FLV 格式在最小化啟動時間方面的效率，TikTok 過去一直使用 Flash Video (FLV) 格式進行直播。然而，FLV 對 CDN 快取造成了挑戰，因為它是連續的、非分段的檔案，使得快取和檢索串流的特定部分變得困難。
- CMAF 作為解決方案：因其分段特性而被認為是 CDN 更具優勢的格式。CMAF 能夠實現更好的快取、可重試性（處理錯誤）和可尋址性（存取特定片段）。
- CMAF 的限制：然而，標準的 CMAF 實作會因為需要在獲取實際的影片和音訊片段之前先獲取資訊清單檔案（其中包含影片和音訊片段的相關資訊）而引入延遲。這些多次請求會增加啟動時間，而啟動時間是 TikTok 的關鍵 QoE 指標。

## 3. CMAF 優化：「Media First」方法

為了應對標準 CMAF 的限制並實現 TikTok 的低延遲目標，TikTok 和 CDN77 合作開發了一系列優化，其中最重要的是「Media First」方法。

- 檔案捆綁 (file bundle)：Media First 的核心思想是將初始播放所需的所有檔案捆綁到一個單元中。這包括資訊清單、影片和音訊初始化檔案，以及第一個影片和音訊片段。
- 減少請求：透過單一請求傳送所有必要的資訊，Media First 消除了多次往返獲取資訊清單和後續片段所造成的延遲。
- 初始播放重點：重點是要注意，Media First 優化主要應用於播放的最初幾秒鐘。在此初始階段之後，將使用標準的 CMAF 傳送機制（資訊清單和片段請求）。

## 4. 挑戰與權衡

CMAF 的實作及其優化帶來了幾個挑戰：

- FLV 效能基準:TikTok 的一個關鍵挑戰是確保優化的 CMAF 效能能夠達到或超過他們現有的 FLV 傳送所實現的效能。多年來, TikTok 在優化 FLV 方面投入了大量精力, 為新的 CMAF 實作建立了很高的基準。
- 低延遲實踐:雖然 CMAF 在技術上能夠實現非常低的點到點延遲(甚至低於 1 秒), 但在 TikTok 的大規模直播平台中實現和維持這種低延遲是一項複雜的工作。長距離資料傳輸和 CDN 基礎架構等因素發揮著至關重要的作用。

## 5. QoE 與效能指標

討論重點介紹了 CMAF 優化對各種 QoE (Quality of Experience, 用戶體驗品質) 和效能指標的影響。

- QoS 指標:TikTok 將其效能指標分為 QoS (Quality of Service, 服務品質) 指標, 包括:
  - 啟動時間:直播開始播放所需的時間。
  - 點到點延遲 (Latency): 從來源到觀眾的整體延遲。
  - 緩衝時間:觀眾遇到緩衝中斷的時間。
- 改進:CMAF 優化顯著改善了以下方面:
  - 減少了啟動時間。
  - 降低了點到點Latency。
  - 由於更有效率的快取和更好的音訊/影片片段同步, 減少了緩衝時間。
  - 透過減少播放所需的請求數量, 優化了行動裝置上的電池使用量。

## 6. 對業務指標的影響

重要的是, 技術優化也對 TikTok 直播平台的關鍵業務指標產生了積極影響。

- 直播滲透率:此指標反映了平台上直播內容的消費量。較低的啟動時間增加了使用者發現和參與直播的機會, 從而提高了直播滲透率。
- 直播觀看時長:此指標衡量使用者觀看直播工作階段的平均時間。減少緩衝 (buffering) 和改善整體 QoE 有助於延長觀看時間。
- A/B 測試:TikTok 在整個優化過程中進行了廣泛的 A/B 測試, 以建立 QoS 指標 (啟動時間、延遲、緩衝) 與這些關鍵業務指標之間的明確關聯。

## 7. 小結

TikTok 和 CDN77 之間的合作展示了持續優化和創新對於大規模提供高品質直播體驗的重要性。Media First 方法顯著提高了 CMAF 效能, 從而增強了 QoS 並對 TikTok 的關鍵業務指標產生了積極影響。對 RTM 和 Media over QUIC 等新協定的持續探索, 突顯了進一步改善直播技術和使用者體驗的承諾。

## 第四章 YouTube與 TikTok在 QoE上的異同之處

### 1. 引言：

數位環境由 OTT(Over-The-Top)串流媒體平台主導，這些平台透過網際網路直接向觀眾傳送影片內容。接下來，以兩個著名的平台：YouTube 和 TikTok 所採用的QoE方法進行比較，以找到未來在串流管理和技術採用等關鍵領域的發展趨勢。雖然兩者都在 OTT 領域運營，但它們的起源、內容重點和使用者參與策略導致了截然不同的技術和營運方法。

### 2.1. YouTube

YouTube 作為線上影片的先驅，已經從一個主要以使用者生成內容 (UGC) 聞名的平台發展成為一個全面的 OTT 服務。其目前的產品包括：UGC 平台：由不同使用者，包括廣播級用戶與創作者用戶同時建立的內容，所形成的大量影片庫。另外還有 YouTube TV，是一種透過網際網路提供傳統線性電視頻道的直播電視服務。該平台上的優質內容，包括：發行電影、節目和直播活動，包括體育賽事。

因此，YouTube 的核心重點包括有：發展規模與可靠性，因為向全球數十億使用者傳送影片。顧及內容多樣性，因為要支援各種內容格式和長度。同時還要擴展營利，經由廣告和訂閱產生各種收入，並且能持續使獲利增長。

### 2.2. TikTok

TikTok 因其對短片行動影片的關注而廣受歡迎。其 OTT 營運的特點是：短片影片，強調簡潔、引人入勝的影片剪輯；直播串流盛行於用戶間，帶來實時(real time)的互動體驗，也越來越重視互動式直播內容。由於主要載體是行動載具，因此必須以行動優先策略為核心，為行動觀看體驗進行優化。

因此，TikTok 的核心重點是：使用者參與最大化，藉此持續吸引使用者互動和在平台上花費更多的時間。UGC 內容探索，透過個人化的動態消息促進相關內容的探索。刺激社群建立，以促進創作者和觀眾之間的互動。因此，TikTok 優先考慮對其使用者體驗至關重要的特定指標，包括有：

- 啟動時間：由於使用者在內容動態消息中滾動的行為，最大限度地縮短直播開始播放的時間是一個主要重點。
- 點到點延遲：實現低延遲至關重要，特別是對於直播購物等互動式直播場景。
- 緩衝時間：減少緩衝中斷對於流暢播放至關重要。



- 電池使用量：考慮到該平台的行動優先特性，優化行動裝置上的電池消耗是一個關鍵考量。

## 2.3 YouTube vs TikTok 兩大平台技術策略比較

| 功能                            | Youtube                                                                                                                                                                     | TikTok                                                                                                                                                                                                                     |
|-------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 核心重點                          | 規模、可靠性、內容多樣性、營利                                                                                                                                                             | 使用者參與、內容探索、社群建立、行動優先                                                                                                                                                                                                       |
| 內容格式                          | 多樣化(UGC、直播、線性、VOD)                                                                                                                                                          | 主要為短片行動影片、不斷成長的直播                                                                                                                                                                                                          |
| 基礎架構                          | 混合式(核心平台 + OTT 擴展)、整合至超級前端節點                                                                                                                                                | 最佳化效能、強調 CDN 整合和用戶端最佳化                                                                                                                                                                                                     |
| 擷取協定                          | 多樣化(RTMP、SRT、HLS、DASH)、探索 Media over QUIC                                                                                                                                   | CMAF(最佳化)、探索 QUIC 和 RTM                                                                                                                                                                                                    |
| QoE 方法                        | 整體、資料驅動、點到點監控、持續改進                                                                                                                                                          | 專注於關鍵指標(啟動時間、延遲、緩衝、電池使用量)、A/B 測試                                                                                                                                                                                           |
| metadata 資料管理                 | 強調標準化和互通性                                                                                                                                                                   | ——                                                                                                                                                                                                                         |
| 技術採用                          | 謹慎的演進、逐步轉型、產業合作                                                                                                                                                             | 敏捷創新、早期採用、效能重點                                                                                                                                                                                                             |
| 延遲優先順序<br>latency<br>priority | 重要，但在其他因素之間取得平衡                                                                                                                                                             | 高優先，特別是對於直播和使用者參與度而言                                                                                                                                                                                                       |
| 延遲最佳化的方法                      | <ul style="list-style-type: none"> <li>* 承認延遲的重要性，特別是對於直播活動。</li> <li>* 根據使用者連線和參與度採用個人化的延遲目標。</li> <li>* 重點是點到點延遲，考量整個傳遞鏈。</li> <li>* 認識到延遲與影片品質和穩定性等其他因素之間的權衡。</li> </ul> | <ul style="list-style-type: none"> <li>* 將低延遲作為其直播平台的核心要求優先考慮。</li> <li>* 特別針對直播購物等互動式使用案例實現低延遲。</li> <li>* 優化 CMAF 傳遞並探索新協定(RTM、Media over QUIC) 以最大限度地減少延遲。</li> <li>* 將啟動時間作為關鍵的延遲相關指標，這是由內容動態消息中的使用者行為所驅動的。</li> </ul> |
| 行動最佳化                         | 重要，但需要適應多樣化的裝置                                                                                                                                                              | 主要重點，驅動許多技術決策                                                                                                                                                                                                              |
| 創新風格                          | 發展現有平台、參與產業標準                                                                                                                                                               | 開發自訂解決方案、嘗試新技術                                                                                                                                                                                                             |
| 使用者互動模式                       | YouTube 多樣化的內容需要更靈活的延遲管理方法                                                                                                                                                  | TikTok 對短片、基於動態消息的消費的關注，需要積極地進行最佳化以實現最小化啟動延遲。                                                                                                                                                                              |

## 2.4 小結：

雖然 YouTube 和 TikTok 都在 OTT 領域運營，但它們截然不同的方法論是由其核心重點、內容策略和使用者參與模式所塑造的。YouTube 優先考慮規模、多樣性和對長期可靠性的關注，而 TikTok 則強調敏捷性、效能和行動最佳化。了解這些差異可以為我們提供有關 OTT 技術不斷發展及其對數位媒體格局影響的寶貴見解。

## 第五章 試論公視+ OTT 平台之 QoE 設計與實踐：公共媒體服務視角

### 1. 引言：QoE 在 OTT 平台的重要性與公共媒體的特殊性

在 OTT (Over-The-Top) 串流媒體服務領域，QoE (Quality of Experience, 體驗品質) 已成為衡量平台成功與否的關鍵指標。它不僅關乎技術層面的影音質量，更涉及用戶在觀看過程中的整體感受與滿意度。對於商業 OTT 平台而言，追求高 QoE 的目標通常與提升用戶黏著度、降低用戶流失率、增加廣告收益等商業目標緊密相連。

然而，公視+作為公共媒體所屬的串流平台，其核心目標除了提供優質的影音內容外，更肩負著傳播公共價值、促進文化交流、服務多元社群等社會責任。因此，以「公視+」的角色，在設計 QoE 指標與執行策略時，必須在商業考量之外，納入公共媒體服務的特殊性，以更全面地評估與提升用戶的觀看體驗。

### 2. 公共媒體服務視角下的 QoE 設計考量

區分「服務品質」(Quality of Service, QoS)和「體驗品質」(Quality of Experience, QoE)兩者間的區別，非常重要。這兩個指標，各自在先前 TikTok 與 YouTube TV 的環節介紹過了。

QoS 通常指網路層級的性能指標，如頻寬、延遲、抖動和封包遺失——這些參數通常由網路營運商管理。相對地，QoE 代表終端使用者對整體串流體驗品質和享受程度的主觀感知。雖然 QoS 因素會影響 QoE，但 QoE 涵蓋了更廣泛、直接影響使用者的元素，包括影片啟動時間、緩衝中斷、影像和聲音品質、播放器穩定性以及易用性。在本章，將特別關注於「公視+」可以控制和提升以直接改善使用者 QoE 的技術能力。

回顧使用者評論和回饋留言，直接證明公視+平台上現存的技術 QoE 挑戰：

- 播放不穩定：使用者回報頻繁的緩衝中斷（「一直轉圈圈」）以及播放中途意外停止的情況（「播一半還會突然停止」）。這些問題直接指向影片傳遞、ABR 邏輯或播放器穩定性方面的問題。
- 播放器功能缺陷：有抱怨指出播放器無法記住使用者上次的播放進度（「記不住之前的播放進度」），導致難以接續觀看。使用者在嘗試快轉或拖動時間軸時也遇到問題，導致進一步緩衝或播放器無響應（「稍微拉一下進度條 就一直轉圈圈」、「選時間軸還會跳掉」）。這表明播放器狀態管理以及串流索引或處理方面存在缺陷。
- 投放問題：使用者對於無法或難以將內容投放到大螢幕表示不滿（「請問何時會支援使用大螢幕播放的功能？」）。這顯示在實施或可靠性方面，投放協定（如 Google Cast 或 AirPlay）可能存在問題。
- 應用程式與相容性問題：有具體錯誤被回報，例如行動應用程式自動旋轉至橫向，但影片卻錯誤地渲染在左下角的一個小區域（「軟體會自動轉成橫向全螢幕，但是影片播放



就是只出線一點，全部在左下角」)。也有使用者指出缺乏對 Android TV 的原生支援，迫使用戶在電視上依賴效果較差的方法，如投放或瀏覽器存取。開發者在一則評論中的回覆也暗示了作業系統更新後可能出現的相容性問題。雖然與不同軟體相關，但政府軟體在桌機上遇到的安裝和權限複雜性突顯了在多樣的使用者環境中確保應用程式順暢運作的普遍挑戰，需要進行全面的審查和優化工作。

前述的是關於OTT串流平台的用戶回饋與技術提升需求。而公視+作為公共媒體所屬的OTT平台，在設計 QoE 指標與執行策略時，除了參考 YouTube 等商業平台的經驗外，更應納入以下公共媒體服務的特殊考量：

## 2.1 內容多樣性與可及性

公共媒體的內容通常具有高度的多樣性，涵蓋新聞、戲劇、紀錄片、教育節目、兒童節目、藝文節目等不同類型，旨在服務不同年齡層、不同興趣、不同文化背景的觀眾。因此，公視+的 QoE 設計應考慮到：

- 不同內容類型的 QoE 需求差異：例如，新聞直播可能更注重低延遲，而紀錄片可能更注重高畫質。
- 不同觀眾群體的 QoE 需求差異：例如，年長觀眾可能對字體大小、操作易用性有更高的要求。
- 內容可及性：確保身心障礙人士也能順利觀看內容，例如提供字幕、手語翻譯、聲音描述等。

## 2.2 公共價值傳播

公共媒體肩負著傳播公共價值、促進社會進步的責任。公視+的 QoE 設計應考慮到：

- 內容的教育性與啟發性：評估內容是否能有效地傳達知識、促進思考、啟發觀眾。
- 內容的社會影響力：評估內容是否能引起社會討論、促進公共參與、推動社會變革。
- 內容的文化代表性：評估內容是否能反映台灣多元的文化風貌、促進文化交流。

## 2.3 服務社群與建立信任

公共媒體旨在服務全體公民，並建立與觀眾之間的信任關係。公視+的 QoE 設計應考慮到：

- 用戶隱私保護：確保用戶的個人資料得到妥善保護，建立用戶對平台的信任。
- 平台透明度：公開平台的運營資訊、內容製作流程等，增加平台的公信力。
- 用戶參與機制：提供多元的管道讓觀眾參與節目製作、提供意見回饋，建立與觀眾的互動。

# 3. 公視+的 QoE 指標設計

基於上述考量，公視+的 QoE 指標設計應包含以下幾個層面：

### 3.1 技術品質指標

這些指標與傳統 QoE 指標類似，主要關注影片播放的流暢性、穩定性和質量：

- 影片啟動成功率：影片成功開始播放的比例。
- 播放失敗率：影片播放過程中發生錯誤而停止的比例。
- 緩衝率：影片播放過程中發生緩衝的頻率和平均緩衝時間。
- 平均位元率：影片播放的平均位元率。
- 啟動延遲：從用戶點擊播放到影片開始播放的時間。
- 延遲變異：啟動延遲隨時間的變化程度。
- 錯誤發生率：播放器發生錯誤的頻率和類型。
- 音訊品質指標：例如音量穩定性、音訊清晰度等。

### 3.2 內容感知指標

這些指標關注內容對用戶的影響，更符合公共媒體的特性：

- 觀看完成率：用戶觀看完整部影片的比例。
- 觀看時間：用戶在不同類型節目上的平均觀看時間。
- 互動率：用戶對影片的互動行為，例如按讚、留言、分享等。
- 內容評分：用戶對影片內容的評分。
- 內容回饋：用戶對影片內容的文字或語音回饋。
- 知識增長：評估教育類節目對用戶知識水平的影響 (例如透過測驗)。
- 情感反應：評估戲劇類節目對用戶情感的影響 (例如透過情緒分析)。
- 文化認同：評估文化類節目對用戶文化認同的影響 (例如透過問卷調查)。

### 3.3 平台使用指標

這些指標關注用戶對平台整體的使用體驗：

- 平台使用時間：用戶在平台上的總使用時間。
- 功能使用率：用戶對平台各項功能的使用頻率。
- 搜尋成功率：用戶成功找到所需內容的比例。
- 導航效率：用戶完成特定任務所需的步驟和時間。
- 介面易用性評分：用戶對平台介面易用性的評分。
- 可及性功能使用率：例如字幕、手語翻譯的使用率。

### 3.4 公共價值指標

這些指標更側重於評估平台對社會的貢獻：

- 多元社群覆蓋率：平台內容觸及不同年齡層、不同興趣、不同文化背景觀眾的比例。
- 公共議題參與度：平台內容引發公眾對公共議題討論的程度。
- 社會影響力指標：平台內容對社會行為、態度、價值觀轉變的影響 (例如透過社會調查)。

- 信任度評分：用戶對平台的信任度評分。
- 品牌形象評估：用戶對公視＋品牌形象的感知。

## 4. 公視＋的 QoE 數據執行方式

### 4.1 數據收集與分析

公視＋需要建立完善的數據收集系統，以獲取上述 QoE 指標數據。數據收集方式包括：

- 用戶端數據收集：在播放器中嵌入程式碼，收集播放相關的數據 (例如緩衝率、啟動延遲)。
- 伺服器端數據收集：收集伺服器端的日誌，記錄用戶的請求和回應 (例如位元率選擇)。
- 用戶行為追蹤：追蹤用戶在平台上的瀏覽、搜尋、互動等行為。
- 用戶回饋收集：提供多元的管道讓用戶提供意見回饋 (例如問卷、評論、客服)。
- 外部數據整合：整合外部的社會調查數據、人口統計數據等，以評估平台對社會的影響。

### 4.2 A/B 測試與實驗設計

進行 A/B 測試和實驗設計，比較不同 QoE 優化策略的效果，並選擇最佳的策略。例如，可以比較不同影片編碼參數對緩衝率的影響，或比較不同介面設計對導航效率的影響。

QoE 優化是一個持續的過程，需要不斷地收集數據、分析數據、調整策略、並再次收集數據，形成一個迭代的循環。

## 5. 公共媒體服務視角下的 QoE 執行策略

儘管公視＋擁有豐富多元、高品質的公共服務內容以及明確的數位轉型目標作為堅實基礎，其技術 QoE 仍面臨顯著挑戰。使用者回饋指出播放穩定性、播放器功能性及跨平台相容性等方面存在問題，這暗示了當前技術基礎設施與應用程式開發方面，存在潛在弱點。此外，現有的數據分析能力較側重於內容表現，而非有效診斷與優化 QoE 所需的精細、客戶端的技術指標。

為應對這些挑戰並提升使用者體驗，應在以下幾個關鍵領域發展核心技術能力：

- 影音傳遞優化：實施先進的內容分發網路(CDN)策略、微調自動調整位元速率(ABR)邏輯、優化編碼工作流程，並為直播串流部署低延遲協定。
- 跨平台播放器強化：提升核心播放器引擎的穩定性與效率、實施健全的錯誤處理機制、確保可靠的播放狀態管理、修復投放(casting)問題，並開發安全的離線播放功能。
- 應用程式穩定性與開發：優先開發原生智慧電視應用程式、提升行動應用程式的性能與穩定性，並簡化使用者介面與驗證流程。



- 品質/格式支援：標準化高品質影像設定檔(HD、4K)、探索 HDR 支援，並確保可靠的多聲道音訊傳遞。
- 數據與個人化基礎設施：建立穩健的使用者設定檔系統，並實施有效且符合倫理的內容推薦引擎。
- 全面性 QoE 監控：部署客戶端監控工具以追蹤關鍵技術性能指標(啟動時間、重新緩衝、錯誤)，並建立數據分析框架以驅動改善。

實施這些優化建議，並著重於基礎穩定性與數據驅動的優化，預期將顯著提升使用者滿意度、增加觀看時長與內容探索、鞏固公視+在台灣 OTT 市場的地位，並最終使該平台能更有效地在數位時代履行其公共服務使命。

應對技術 QoE 挑戰並實施建議的能力需要一個具有明確優先順序的策略性方法。下表整理了多項建議，並提出優先順序框架：

| 建議領域        | 具體行動                                        | 影響的關鍵 QoE 指標             | 依賴條件           | 建議優先級 |
|-------------|---------------------------------------------|--------------------------|----------------|-------|
| QoE 監控與分析   | 實施客戶端監控 SDK; 建立數據聚合/分析平台; 配置警報; 建構 A/B 測試框架 | 能夠衡量所有其他指標               | 技術資源           | 高     |
| 播放器/應用程式穩定性 | 修復核心播放錯誤、緩衝原因、崩潰、渲染錯誤; 改善錯誤處理/恢復            | 重新緩衝比率、播放錯誤、啟動時間         | QoE 監控數據       | 高     |
| 播放器功能性      | 修復播放進度保存; 改善搜尋可靠性; 穩定投放功能                   | 使用者滿意度、會話持續時間            | 播放器核心優化        | 高     |
| 原生智慧電視應用程式  | 開發/發布適用於關鍵電視平台的應用程式 (Android TV, tvOS 等)    | 觸及率、使用者滿意度 (電視觀眾)、會話持續時間 | 穩定的播放器核心、技術資源  | 中-高   |
| 影片傳遞優化      | 調校 ABR 邏輯; 優化編碼; 實施多 CDN (可選); 部署低延遲 (直播)   | 重新緩衝比率、影片品質、啟動時間、延遲 (直播) | QoE 監控數據、播放器增強 | 中     |

|         |                                   |                     |                    |     |
|---------|-----------------------------------|---------------------|--------------------|-----|
| 離線觀看    | 開發安全的下載/播放系統                      | 功能可用性、使用者滿意度 (特定群體) | 穩定的播放器/應用程式、DRM 整合 | 中   |
| 品質與格式支援 | 標準化 HD/4K 設定檔; 實施 HDR/多聲道音訊 (若適用) | 影片品質、音訊品質           | 編碼基礎設施、播放器能力       | 中-低 |
| 數據與個人化  | 建立使用者設定檔/歷史系統; 實施/增強推薦引擎          | 內容探索、參與度、會話持續時間     | 後端基礎設施、數據隱私框架      | 中-低 |

### 5.1 透明化 QoE 評估

公開 QoE 評估方法和指標，增加平台的公信力，並讓觀眾了解平台在提升用戶體驗方面的努力。

以公視+ 這樣以廣泛服務公眾並在日益數位化的媒體環境中保持相關性為使命的平台而言，提供高品質的技術體驗不僅是競爭優勢，更是基本要求。無縫、可靠且高保真的觀看體驗能確保公視製作或策劃的寶貴公共內容能夠順暢地觸及目標觀眾。反之，糟糕的技術 QoE——例如載入時間過長、頻繁緩衝、播放錯誤或應用程式崩潰——會導致使用者感到沮喪、放棄觀看，並最終阻礙平台有效履行其公共服務使命。因此，透過卓越的技術來確保可及性和正面的使用者體驗，對於實現公視+ 的策略目標至關重要。

### 5.2 重視多元社群的 QoE 需求

針對不同年齡層、不同背景的觀眾，設計不同的 QoE 優化策略。例如，針對年長觀眾，可以提供更大的字體、更簡化的操作介面；針對身心障礙人士，應確保平台的可及性。

### 5.3 強化內容品質與公共價值

除了技術層面的 QoE 優化，更應重視內容的品質與公共價值。提供高品質、有深度、有啟發性的內容，才能真正提升用戶的觀看體驗和對平台的認同。

### 5.4 建立雙向互動機制

提供多元的管道讓觀眾參與節目製作、提供意見回饋，並積極回應觀眾的需求。這不僅能提升用戶的參與感，也能幫助平台更準確地了解用戶的 QoE 需求。

### 5.5 合作與創新

與學術界、產業夥伴合作，共同研究 QoE 評估方法和技術，並探索新的 QoE 優化策略。例如，可以研究如何利用人工智慧技術，提供更個人化的 QoE 優化。

## 6. 小結

當代數位媒體生態系統中，OTT 平台提供的技術品質參與度與其成功息息相關。對於公視＋而言，實現技術卓越不僅僅是為了與商業服務競爭；它是在 21 世紀有效履行其公共服務使命的基礎。高 QoE 能培養使用者信任、鼓勵更深入地參與有價值的公共內容、確保廣泛的可及性，並最終加強台灣公共廣播機構與其觀眾之間的聯繫。QoE 是 OTT 平台成功的關鍵，對於公共媒體所屬的串流平台「公視＋」而言，QoE 的設計更應納入公共媒體服務的特殊性。

透過優先考慮基礎穩定性、實施徹底的監控，並採用迭代方法來增強其技術能力，公視＋可以克服當前的挑戰。對技術卓越的持續投資，與其獨特的內容優勢和公共服務價值觀<sup>11</sup>相結合，不僅對於滿足現有使用者是必要的，對於吸引新觀眾並鞏固公視＋作為領先、值得信賴且不可或缺的台灣 OTT 平台的地位也至關重要。利用技術改進是釋放公視＋高品質、多元化且具有文化意義的內容庫全部潛力，造福所有觀眾的關鍵。

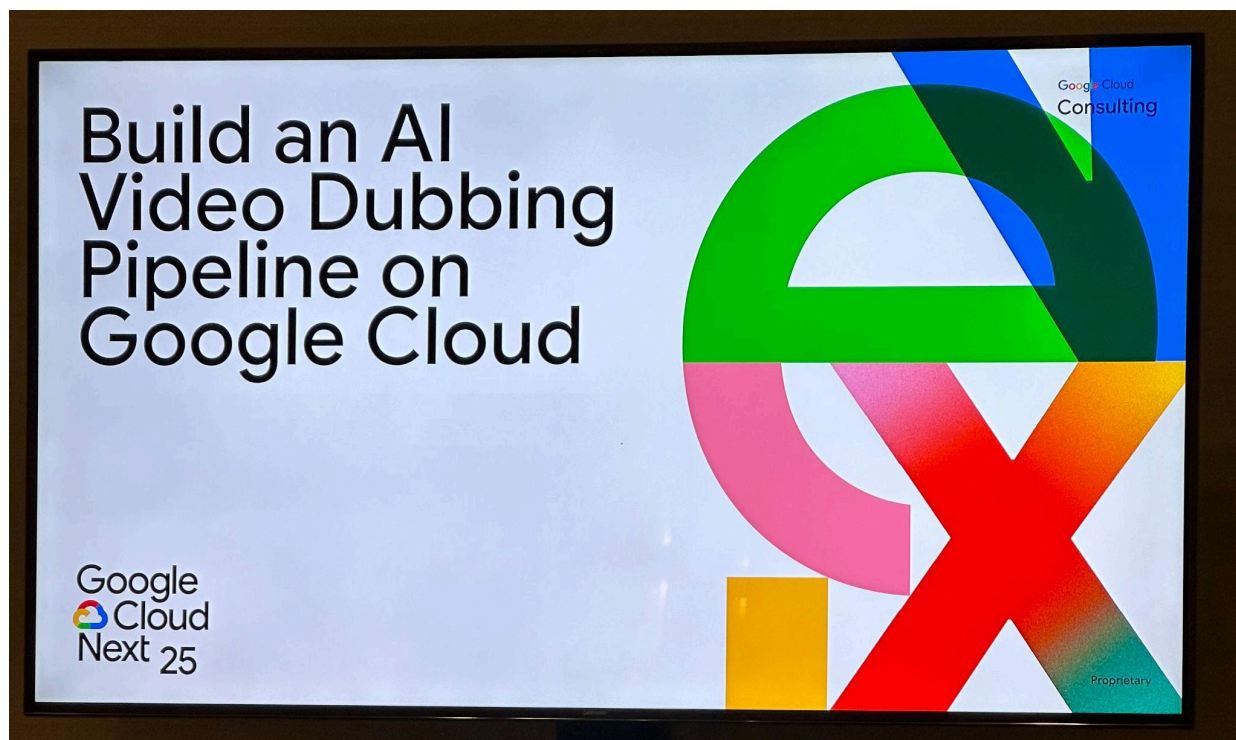
公視＋應建立一個全面、客觀、且具有公共價值導向的 QoE 評估體系，並透過透明化、多元化、互動化、合作化等策略，持續提升用戶的觀看體驗，實現公共媒體的社會責任。



## 第二部分

### 第六章：美國串流平台關於 Dubbing AI 的快速實踐

關於人工智能驅動的配音技術分析



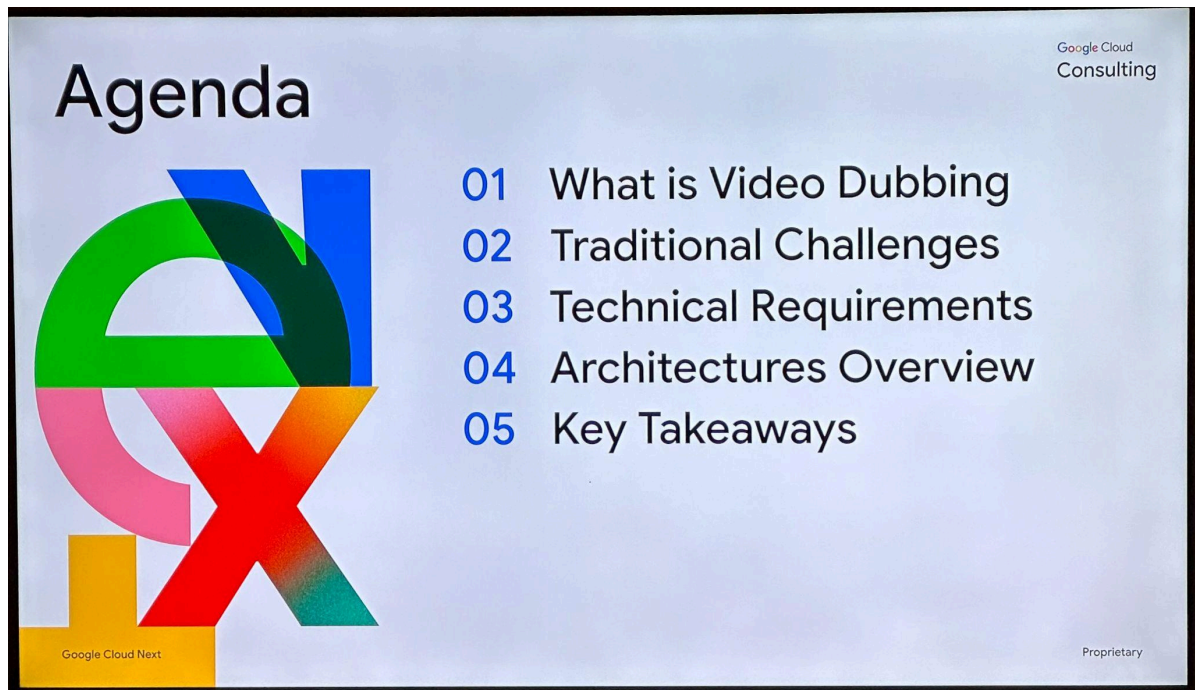
#### 1. 引言

配音AI是一種可將影片翻譯並配音成多種語言的技術，使內容更容易被全球觀眾所接受。它的目標是打破語言障礙，讓原本使用某語言的影片能自然地用另一種語言呈現。不單只是提供了多語字幕，更提供了多語對嘴，使發音與嘴型得以同步。

目前配音AI仍面臨許多挑戰，包括高昂的成本、人力與計算資源需求，以及對翻譯準確性與同步的要求。

有效的配音AI需具備以下條件：準確性、品質、文化細節處理、快速處理時間、可擴展性、成本效益、可自訂性、可控性與合規性。

本部分以Google最新推出的工具「Dubbing AI」為主體，分析了人工智能(AI)在影片配音中的應用，討論了一個旨在簡化和增強配音過程的平台。本分析將涵蓋傳統配音中固有的挑戰、人工智能驅動解決方案的技術要求、人工智能驅動的配音管道所涉及的工作流程，以及實施的各種考量。

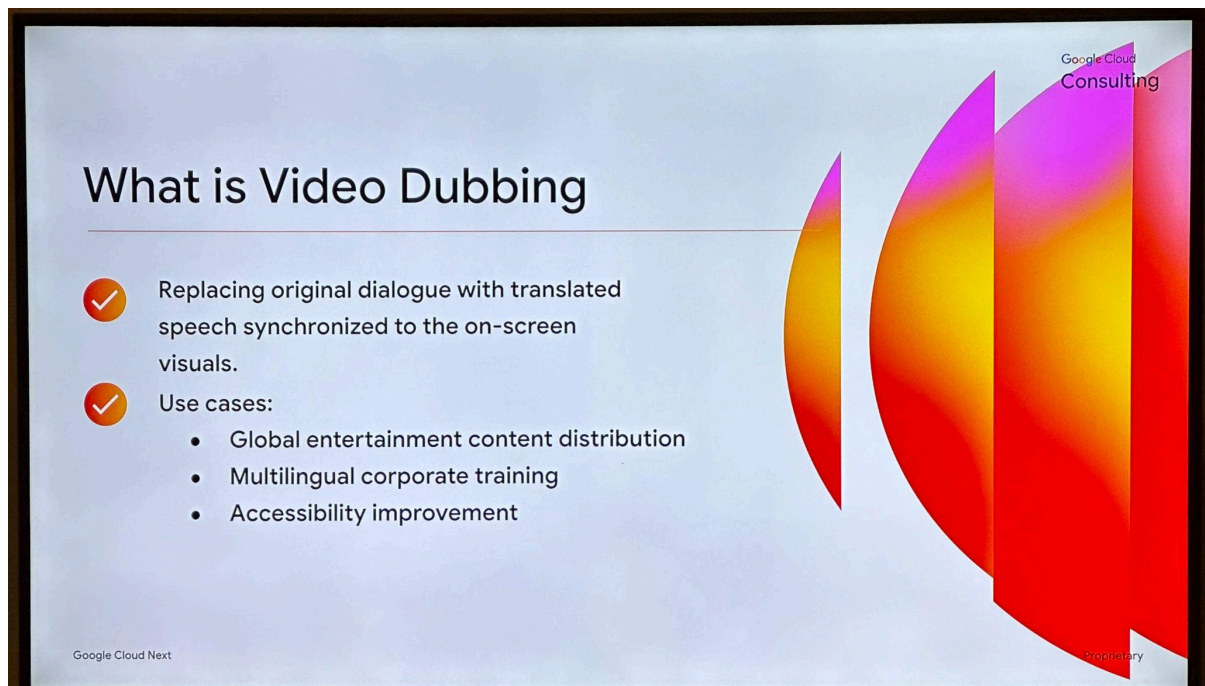


## 2. 傳統配音的挑戰

傳統影片配音，即用不同語言的翻譯版本替換原始音軌的過程，被認為是一個複雜且昂貴的努力。

### 2.1 起心動念

在美國市場當然是說英語。但實際上，放大到全球市場，只有不到20%的人理解並使用英語。為了讓更廣泛的受眾能輕鬆省力看懂非自身母語的影片，如果只想要同一個影片，能自然地說另一種語言，是這個專案的目標。配音用的人工智慧，可以接觸到更多的受眾。一切皆有可能。

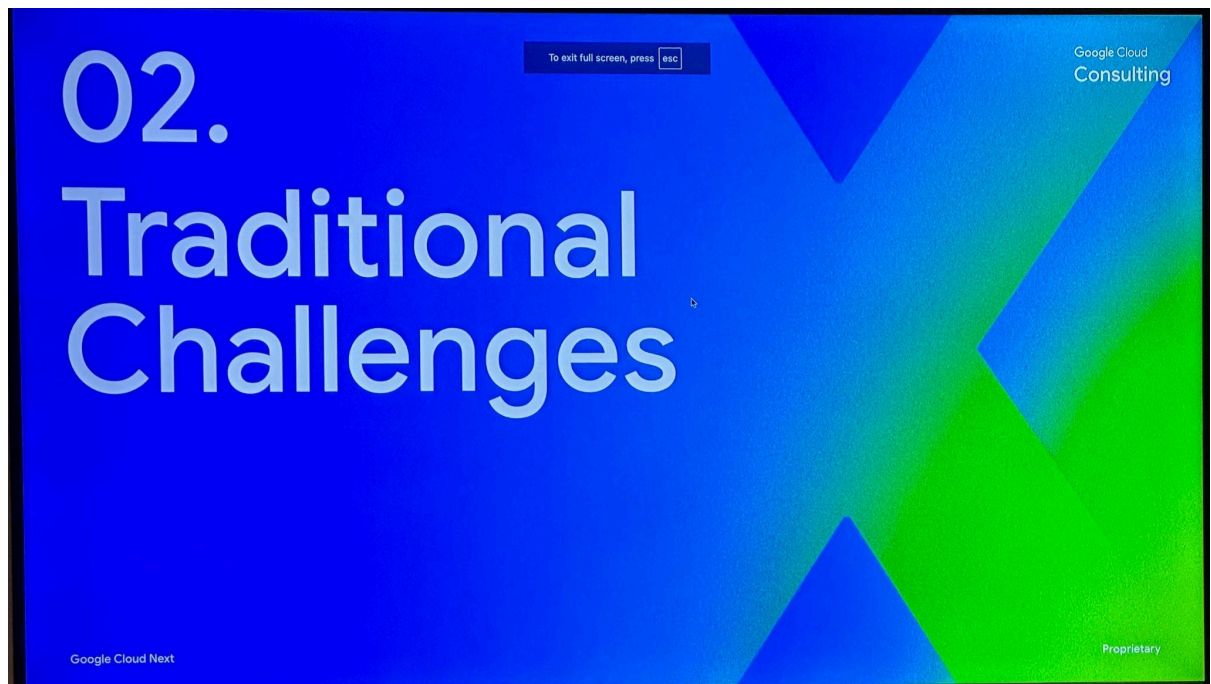


但現實中存在著許多挑戰。首先，這是一種非常昂貴的程序：一是勞力密集，另一個是計算密集，然後是金錢密集。

首先是大量的後端人工處理，如果在特定的目標語言中，做一個一分鐘的後端處理，通常要花費幾十美元。這流程很慢，需要經過很多輪的編輯與校正，才能使文字、聲音和影片都正確且同步。而配音方面的挑戰，除了獲得源語言素材，接著需要將其翻譯成目標語言，而且要確保正確。前面這個部分是屬於字幕。接著是關於配音。

目標是能找到類似原始角色的聲音，或是聲音複製，以及聲音背後的所有情感，希望盡可能多地將其傳達給目標語言。然後是口型同步。如果只是閱讀字幕而沒有口型動作，那是一種不同的體驗。觀看影片時，希望能夠專注於視覺和故事本身。



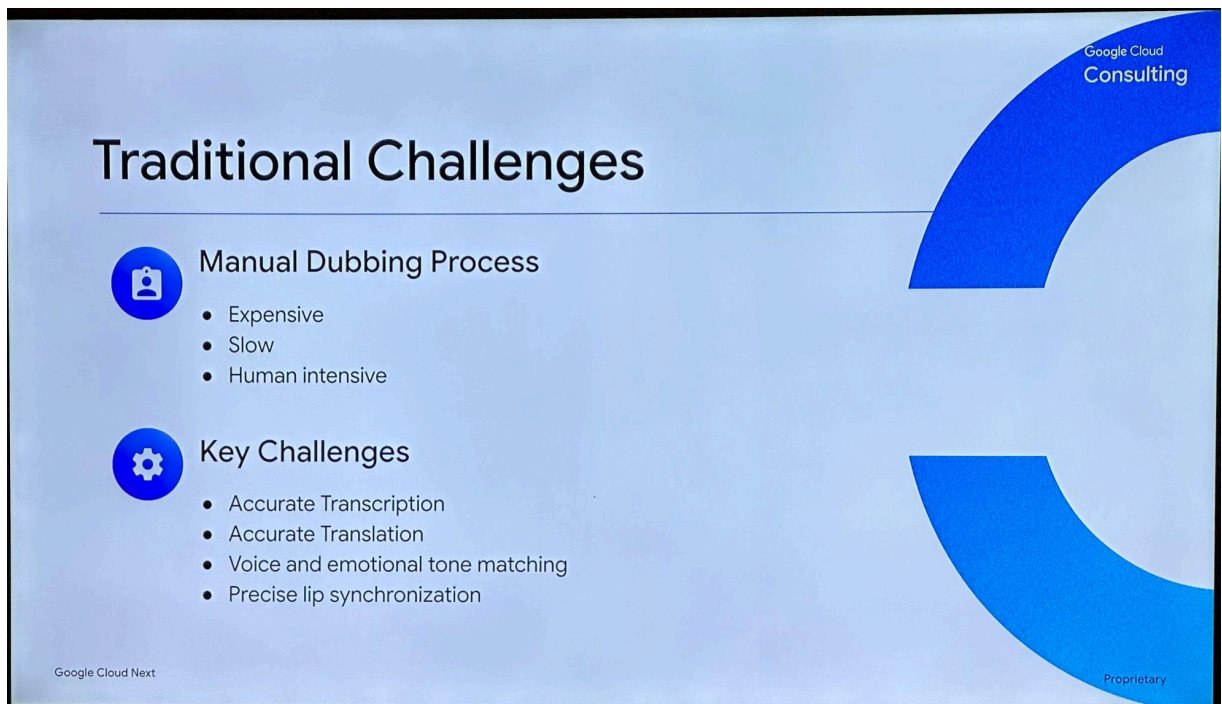


## 2.2 技術挑戰

如果要應對這些挑戰，就要考慮哪些技術能實現？

首先，是準確性和品質。文字記錄的準確性，對輸出品的品質影響最大。所以文化細微差別，例如，即使人工翻譯，語言在不同地區，使用不同的俚語，或不同的習慣用語，當傳達這些並使其真正本地化內容時，並不容易。第三，如果是要為一個小時的評論進行配音，且使用傳統的方式，可能需要數週甚至更長的時間。

這些傳統的流程，是否有可能因為科技而縮短？或減少成本耗費？以下是影片若要做多語發行，在傳統流程中會會面臨到的各種挑戰，包括有：



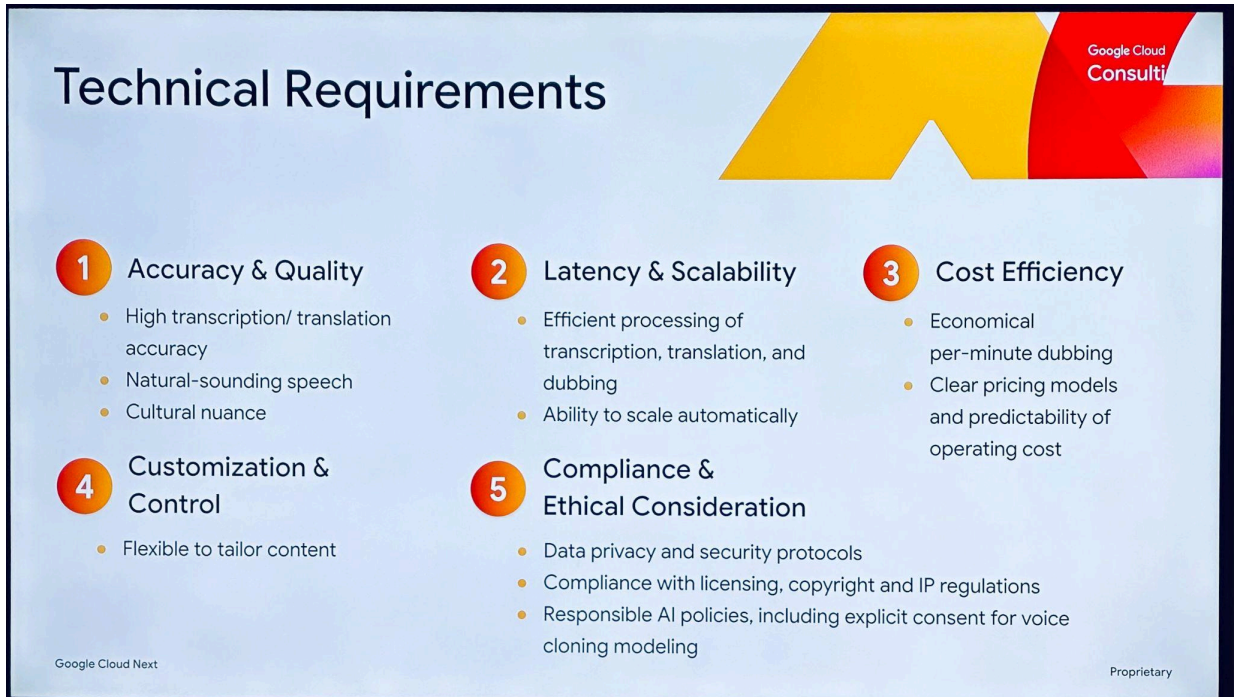
成本：與傳統配音相關的巨額成本。且此成本是多方面的：

- 勞力密集：配音需要大量的人力參與，從翻譯和改編到配音和同步。這意味著高昂的人力成本。
- 計算密集：配音中，音頻和視頻處理所需的計算能力可能非常大，增加了整體費用。
- 金錢支出：勞動力和計算資源的結合使配音成為一項資金密集型的過程。文本明確指出，即使是短時間的配音內容（例如，一分鐘）也會產生相當可觀的成本。
- 時間：傳統配音是一個緩慢的過程，通常涉及多輪編輯和修訂，以確保準確性和品質。這種漫長的周轉時間對於需要快速接觸全球受眾的內容創作者和發行商來說可能是一個瓶頸。
- 技術難題：過程中，有幾個導致配音複雜性的技術挑戰：
  - 源材料分析：準確轉錄原始音頻，包括識別說話者和時間，是至關重要但困難的第一步。
  - 翻譯和本地化：在保留文化細微差別、俚語和習語的同時將源語言翻譯成目標語言，是一個複雜的語言挑戰。一種語言內的不同方言和地區差異進一步使這個過程複雜化。
  - 聲音複製和情感傳遞：複製原始說話者的聲音並在目標語言中傳達適當的情感，在技術和藝術上都是具有挑戰性的。

- 口型同步：在配音音頻和演員的嘴唇動作之間實現準確的口型同步，對於自然和可信的觀看體驗至關重要，但需要精確的音頻和視頻處理。

### 3. 「AI配音」解決方案的技術要求

這項專案確定了人工智能驅動的配音解決方案，必須滿足的幾個關鍵技術要求，以應對傳統配音的挑戰並提供可行的替代方案。



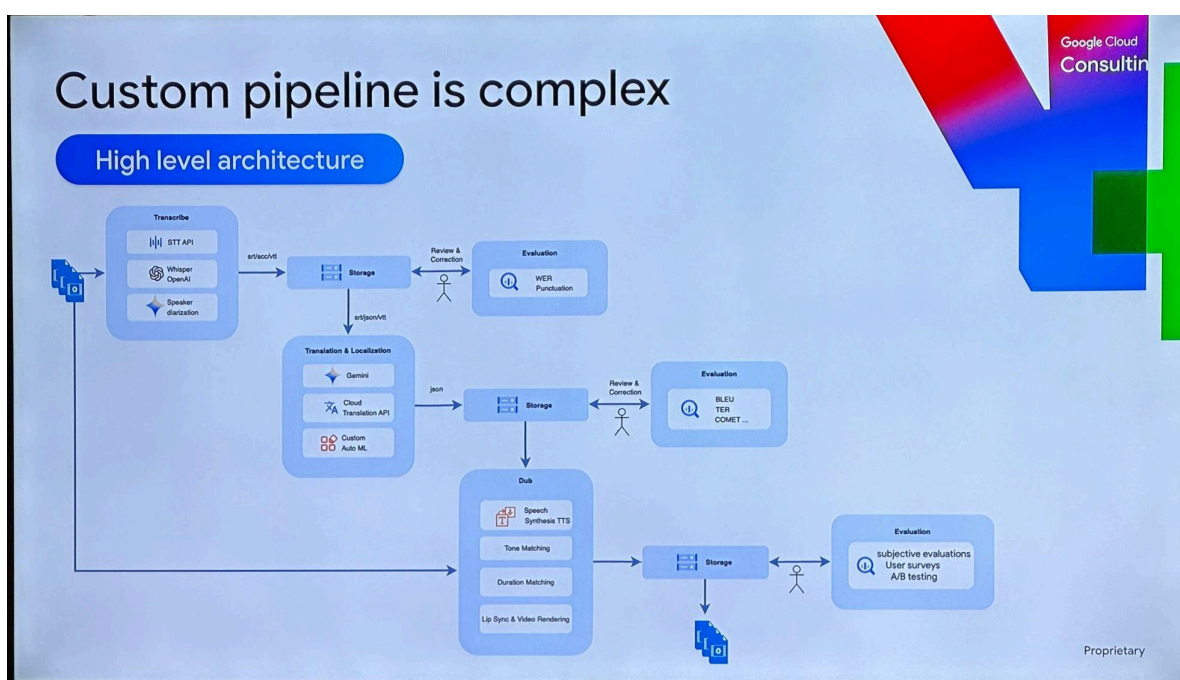
- 準確性和品質：
  - 轉錄準確性：初始轉錄的準確性至關重要。轉錄中的錯誤將會貫穿整個管道，對最終配音產品的品質產生負面影響。
  - 翻譯準確性和細微差別：人工智能必須準確翻譯源語言，同時保留文化細微差別、俚語和習語。這需要複雜的自然語言處理(NLP)模型，能夠理解並適應不同的語言環境。
  - 音頻品質：生成的音頻必須清晰、自然、沒有雜音。
- 速度和周轉時間：人工智能驅動的解決方案應顯著縮短配音的周轉時間，實現更快的內容本地化和分發。目標是從數週縮短到數天、數小時，甚至即時配音。
- 可擴展性：該解決方案必須能夠有效地擴展，以處理大量內容並支持多種目標語言，而無需重大的基礎設施變更。
- 成本效益：人工智能驅動的配音應該比傳統方法更具成本效益，降低勞動力和計算成本。
- 客製化和控制：該平台應為用戶提供對配音過程的控制程度，允許他們微調參數並在必要時進行干預，以確保高品質。



- 合規性：
  - 聲音複製的同意：獲得聲音複製的適當同意對於解決法律和道德問題至關重要。
  - 數據隱私和安全：保護數據源、中間結果和最終輸出的隱私和安全。
  - 機器可檢測性：遵守新興的標準和法規，例如歐盟對機器可檢測的人工智能生成內容的要求，至關重要。

#### 4. 「AI配音」工作流程

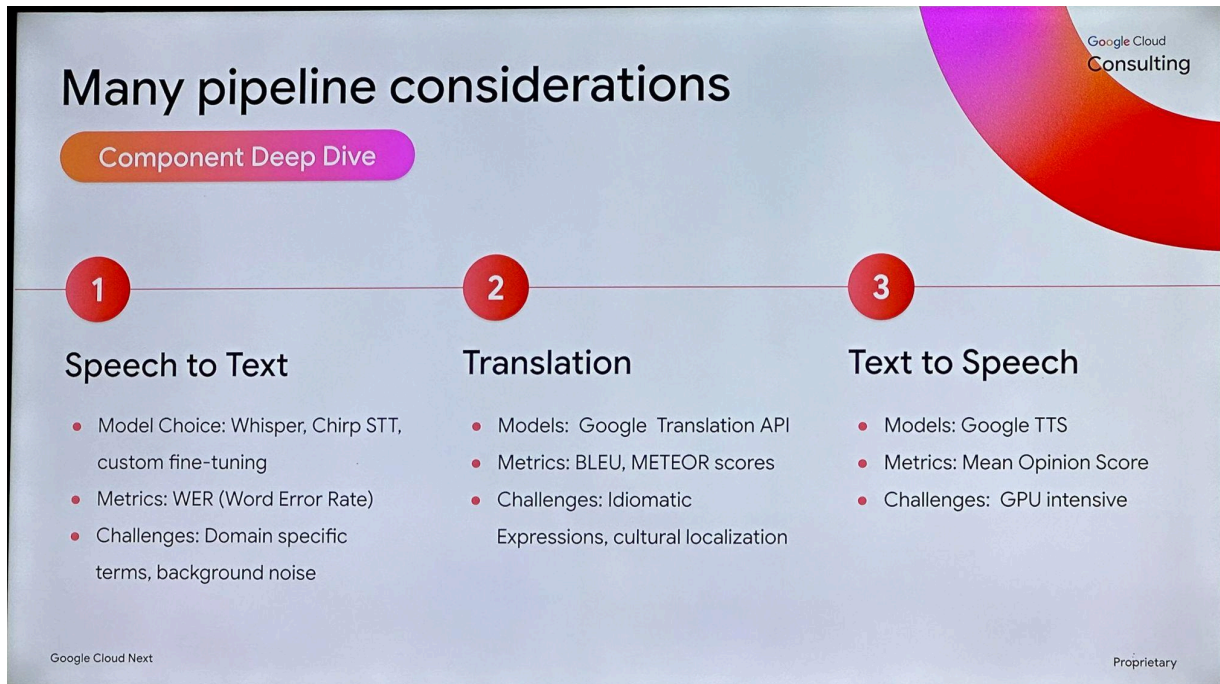
接下來的工作流程圖，描述了人工智能驅動的配音管道的一般工作流程，強調了所涉及的關鍵階段。這流程也的確頗為複雜。



圖說：AI 同步配音工作流程圖

- 音頻處理：初始階段涉及處理源音頻。這可能包括：
  - 音頻分離：將語音與背景噪音、音樂和其他聲音分離。這對於準確的轉錄至關重要。所以真正施作的時候，建議是要能夠分離背景音，盡量讓對白清晰可辨。其實無論是傳統方式（涉及人工聽打）或採取人工智能方式，很多時候的失誤都是出在這一段：由於環境噪音過大，無法辨識正確內容，導致出錯。
  - 轉錄：分離的音頻被轉錄成文字（voice-to-text）。理想情況下，轉錄包括時間戳記和說話者識別，以方便後續步驟。
- 評估：評估轉錄的準確性。這可能涉及：
  - 自動評估：使用指標和算法來評估轉錄的品質。
  - 人工審查：讓人工審查員糾正錯誤並確保準確性，特別是對於複雜或細微差別的內容。人工校對在判定內容準確度的涉入，也是最重要的。

- 翻譯和本地化：準確的轉錄被翻譯成目標語言。此階段可能涉及：
  - 通用翻譯模型：使用通用翻譯API。
  - 客製化模型：使用專門或客製化的模型來處理特定的術語、方言或文化背景。



- 文字轉語音 (**TTS**)：翻譯後的文本，使用人工智能驅動的TTS技術轉換為語音。此階段需要：
  - 聲音複製：複製原始說話者的聲音。前面有講過，這部分需要授權同意，無論是聲音的主人，或是內容發行商。
  - 情感語調：生成傳達適當情感的語音。
- 聲音複製和音調調整：此階段著重於細化生成的音頻，確保聲音被準確複製並反映正確的音調、韻律和其他聲音特徵。
- 音頻和視頻混合：生成的音頻與原始視頻混合，包括口型同步。這可能涉及：
  - 音頻驅動的面部重現動畫：視需要修改視頻影格，以完美匹配配音音頻。
  - 音頻對齊：將配音音頻與現有的嘴唇動作對齊。
- 人工參與的審查：在流程的各個階段可能需要人工審查員參與校正，以確保品質並進行必要的調整。
- 最終輸出：生成最終的配音視頻。



# Many pipeline considerations

Component Deep Dive

Google Cloud Consulting

4

## Voice Cloning & Tone Matching

- Techniques: Cross lingual voice transfer, multi-speaker fine-tuning
- Evaluation: Speaker similarity embeddings, human testing

5

## Lip-sync modeling

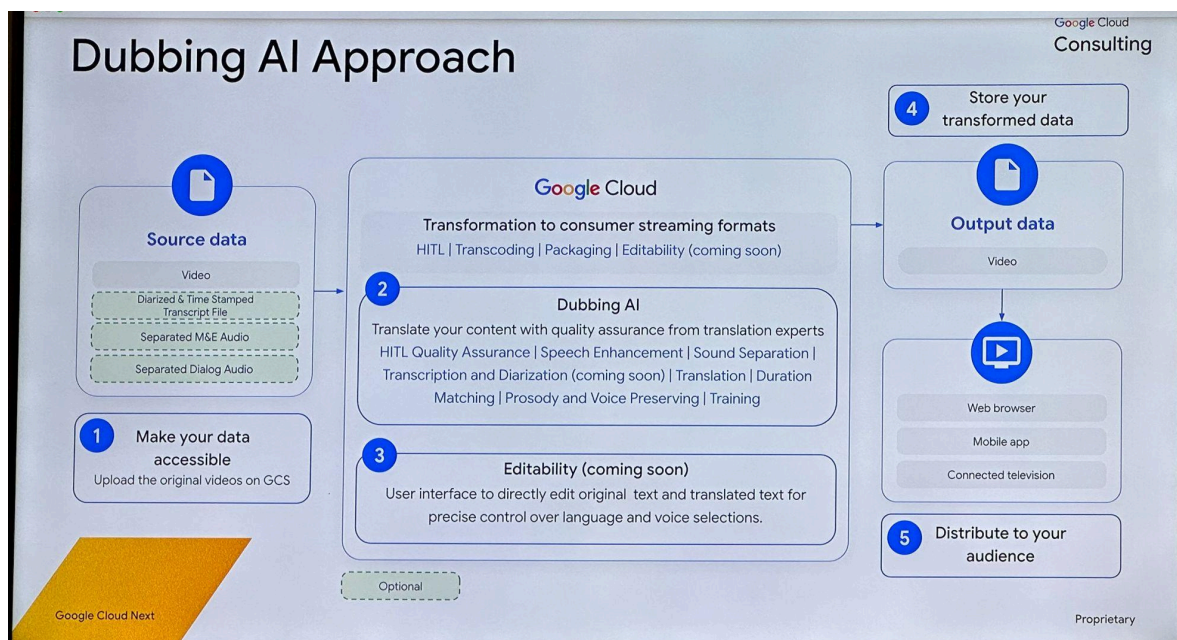
- Models: Wav2Lip, Google Proprietary
- Methods: Audio-driven face reanimation, Audio alignment
- Challenges: GPU requirements, Ethical considerations, Personal data

Google Cloud Next

Proprietary

## 5. 實施方法：客製化與託管平台

專案過程中，需要建立一個流程評估框架。執行專案的過程，也需要正確設定指標。如果評估狀態很糟糕，例如成果低於某個閾值，就需要以此作為依據來更正它，然後再發回去 系統重新執行，或者只是新增某些輸出條件。實施人工智能驅動的配音的兩種主要方法：



圖說：AI 智慧同步配音方法論

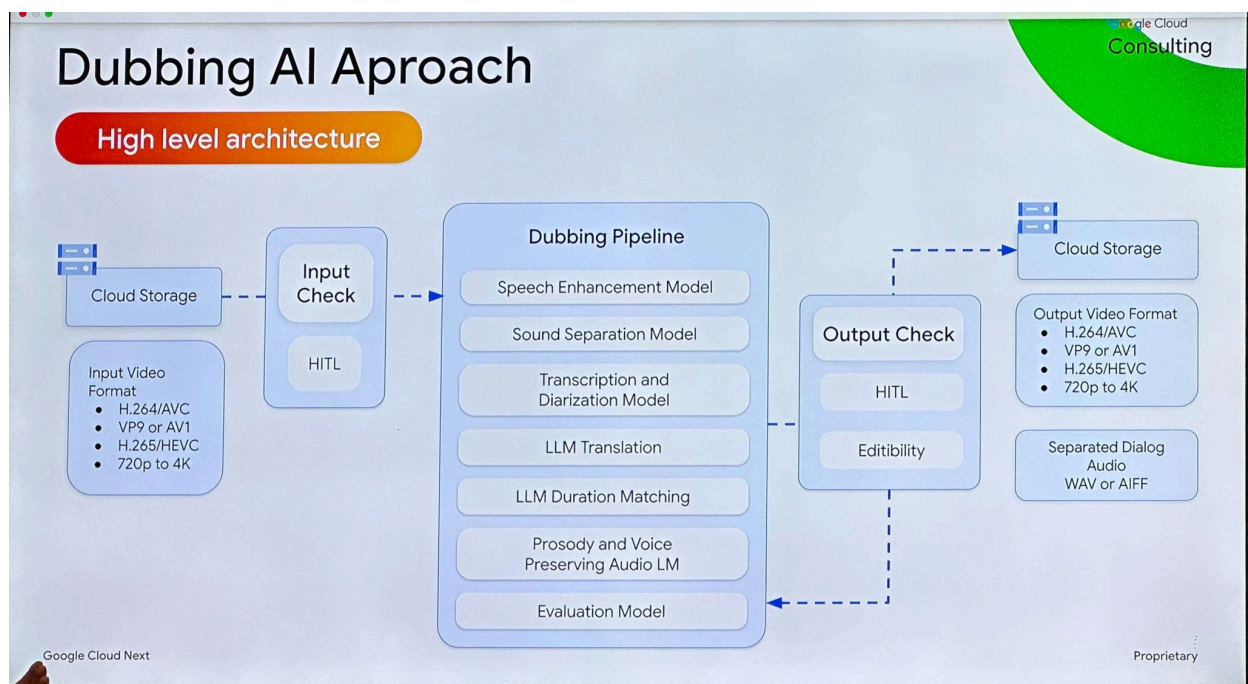
- 客製化構建管道：
  - 控制和靈活性：這種方法提供了最大的控制和靈活性，允許用戶客製化管道的每個組件。



- 複雜性和責任：用戶負責管理整個工作流程的複雜性，包括整合、擴展、維護和合規性。而且，每當涉及到複製時，要把所有的合規性銘記在心。
- 資源密集型：需要對基礎設施、工程資源和專業知識進行大量投資。
- 託管平台：
  - 簡化工作流程：託管平台封裝了大部分複雜性，提供了更簡化和用戶友好的體驗。
  - 責任減少：平台提供商處理許多技術挑戰和合規性問題。
  - 隨用隨付：通常提供隨用隨付的定價模式，降低了前期成本。

這些方法之間的選擇取決於諸如用戶的資源、技術專業知識、長期目標以及其配音需求的具體要求等因素。

Google Translation API有大量的訓練資料。即使對於低資源語言，它也有非常好的結果。但對於一些高階公司來說，他們想要客製化一個模型（例如廣東話，或是印度的某種方言，又或是亞美尼亞語等，或是一種較少人用的語言）。因為這種語言，有一些特定的詞語，使用者想要更好地處理它。有一些特定的主題。甚至想要處理一個特定的語料庫。



圖說：AI 智慧同步配音方法論

## 6. 關鍵技術考量和挑戰

AI 配音中的幾個關鍵技術考量和持續存在的挑戰。像是翻譯，是需要妥善處理的另一項工作。對於很多像新聞這樣的東西，比如訓練材料，這些都比較容易處理，但當涉及到像特定的文化時。當翻譯結果呈現出來時，通常會使用其中一些指標來評估翻譯結果是否達成期望值。

- **準確的轉錄:**實現高度準確的轉錄,特別是在存在背景噪音、音樂和重疊語音的情況下,仍然是一個重大的挑戰。
- **情境翻譯:**如前所述,準確翻譯嚴重依賴文化背景、習語和俚語的語言對於人工智能模型來說是困難的。但目前我們的最終目標不是,比如,翻譯到最高的準確度,而是符合情境的轉譯。例如當英語轉譯成德語,整體句型會很不同,甚至會變成長單字。因此人工審閱部分對於翻譯來說至關重要。
- **口型同步精度:**實現完美的口型同步,特別是通過音頻驅動的面部重現動畫,在技術上是複雜且計算成本高昂的。像是唇形同步模型,市面上雖說也有一些流行的模型,而Google也有自己的研究模型。兩種方法各有不同。但若是關於真正的唇形重製,就像它與角色說出的詞語完全匹配一樣。還需要做音訊驅動的面部重製動畫,這意味著我們修改影片幀,使其完美。
- **計算資源:**人工智能驅動的配音,特別是聲音複製和口型同步等任務,可能是計算密集型的,需要大量的GPU資源。
- **評估指標:**開發強大而可靠的評估指標來評估配音管道每個階段的品質對於品質控制至關重要。
- **人工參與的整合:**確定工作流程中人工干預的最佳時間點

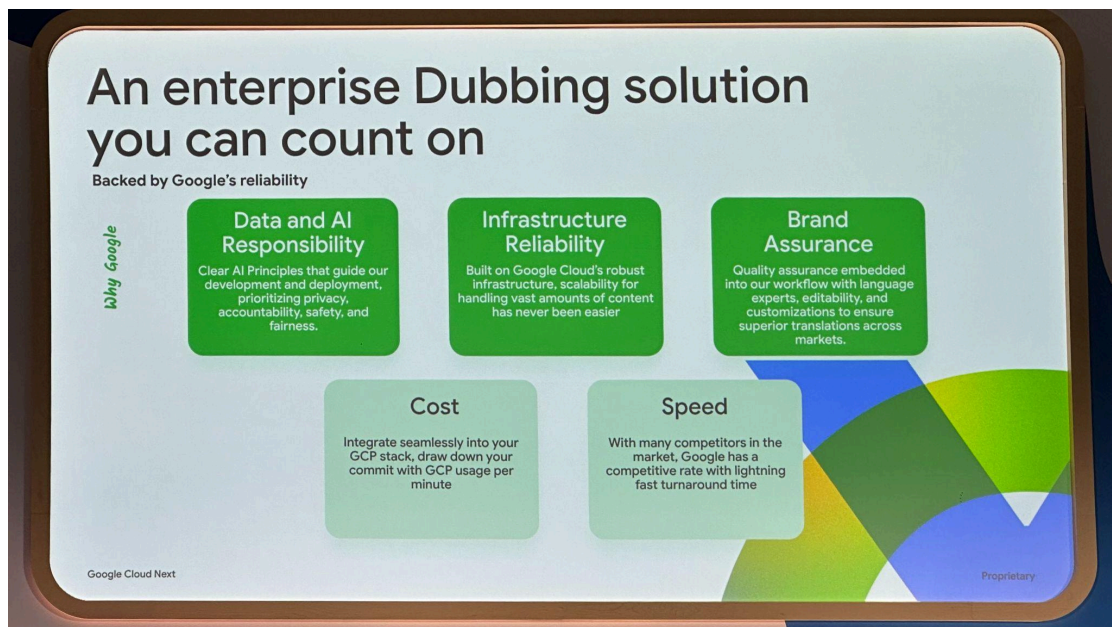
圖說:AI 智慧同步配音 分階段工作流程圖

Max是華納兄弟探索提供的串流媒體影片服務平台，其前身HBO Max。MAX 平台的使命是提供使用者一個最愛的串流選擇，內容涵蓋娛樂、新聞、體育與多元文化節目。為提升用戶體驗，團隊運用人工智慧與機器學習來實現個人化內容推薦，讓用戶能更快找到他們喜歡的節目，同時減少介面阻礙，創造沉浸式體驗。





其中，MAX 推出 AI 自動產生字幕功能『Caption AI』，與 Google 合作開發，將字幕生成與位置最佳化自動化，大幅降低 50% 成本並提升全球內容可及性，支援 40,000+ 部作品與多語言版本。



而在其他的演算法運用上，MAX 引入「整頁最佳化」功能，不只考慮內容排序，還能因應用戶偏好動態調整展示位置，並透過顯性與隱性的用戶訊號（如觀看歷史、點讚、跳過）不斷優化推薦結果。



## 8. 總結

AI 配音技術正處於一個快速發展的階段，它將改變影片內容本地化和全球傳播的方式。

### 傳統配音的困境與 AI 的解決方案

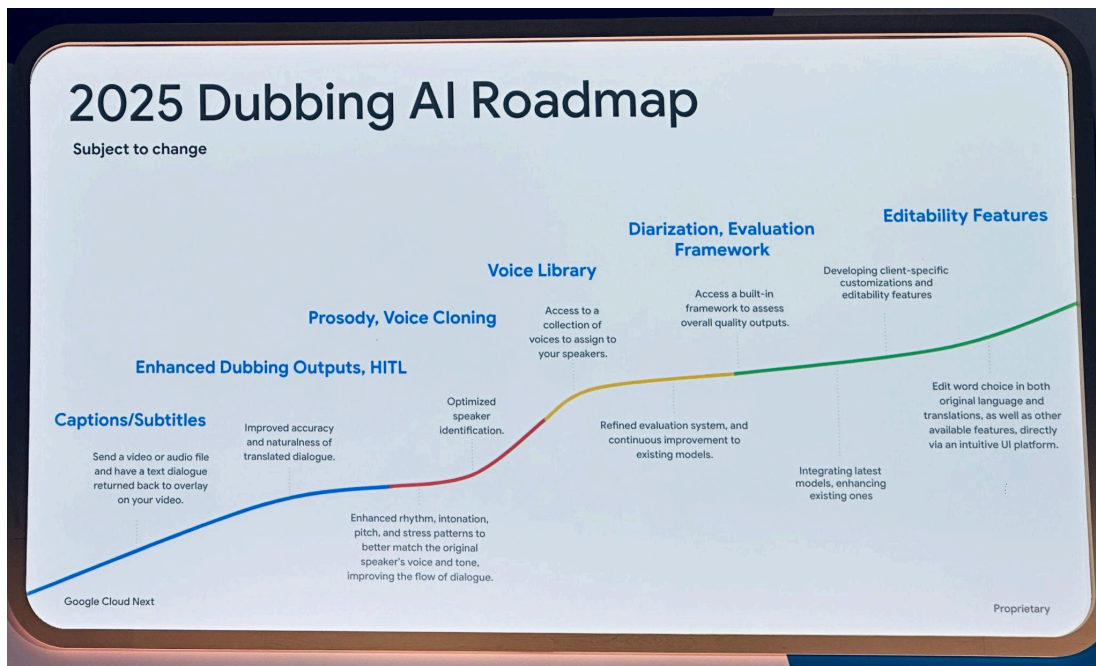
傳統配音一直以來都面臨著成本高昂、耗時長、技術難度大等問題。高昂的人力成本、複雜的音視頻處理需求，以及對翻譯、本地化和口型同步的高標準要求，都使得傳統配音成為一個耗費大量資源的過程。此外，確保情感的準確傳達和聲音的自然度也是一大挑戰。

AI 配音為這些問題提供了潛在的解決方案。透過自動化的音頻處理、轉錄、翻譯和語音合成，AI 能夠顯著降低成本並縮短製作時間。精確的轉錄和翻譯有助於確保內容的準確性，而先進的文本轉語音 (TTS) 技術和聲音複製則能夠在一定程度上還原原始聲音的情感和特徵。此外，AI 還能夠處理大規模的內容，並支持多種語言，實現快速的全球發布。

### 技術挑戰與倫理考量

為實現高品質的 AI 配音，需要滿足一系列的技術要求。轉錄和翻譯的準確性至關重要。生成的音頻必須清晰、自然且沒有雜音。AI 解決方案還需要具備快速、可擴展和成本效益高的特點。

儘管 AI 配音技術具有巨大的潛力，但仍面臨著一些挑戰。準確的轉錄、情境翻譯、口型同步精度、計算資源需求，以及開發可靠的評估指標都是技術上的難題。此外，如何平衡自動化與人工干預，確保高品質的輸出，也是一個需要仔細考量的問題。倫理和法律考量同樣重要。聲音複製的同意、數據隱私和安全，以及符合相關法規（如歐盟對 AI 生成內容的機器可檢測性要求）都是必須解決的問題。



## 未來的發展方向

AI 驅動的配音技術仍在不斷發展。AI 模型的持續改進將帶來更準確、更自然、更細膩的配音效果。即時配音的實現將為現場直播和互動內容帶來新的可能性。個人化配音體驗、自動化工作流程優化，以及更完善的評估指標和倫理規範，都將是未來的發展方向。透過解決傳統配音的挑戰，AI 能夠實現更快、更具成本效益和更易於訪問的內容分發。然而，要實現這一潛力，需要不斷的創新和對技術、倫理和法律複雜性的仔細考量。開發強大的人工智能模型、高效的工作流程和完善的倫理指導方針，對於這種變革性技術的成功實施和廣泛採用至關重要。AI 配音技術的發展前景廣闊，但同時也面臨著許多挑戰。只有透過不斷的努力和合作，才能充分發揮其潛力，為全球內容創作者和觀眾帶來更好的體驗。