

一、指導老師：吳世弘

二、組員：陳禹安 (11127010)、楊祖威 (11127090)

三、系統環境：

軟體：作業系統 (Linux (Ubuntu 22.04)、Windows 10)、開發與運行 (Python、VSCode、Unity)、Ollama_LLM、Whisper_ASR、SoVITS_TTS、YOLO Object Detection、xArm_Python_SDK

硬體：遠端伺服主機 x2 (GPU：NVIDIA RTX 4090 / RTX 5090)、本地端控制終端 Notebook x2、xARM6、Intel RealSense D435i、音響與顯示器設備

四、系統功能與特色：

(一)功能

1. 語意理解與情境判斷

系統透過大型語言模型 (LLM) 理解使用者輸入，判斷其需求屬於聊天互動或調酒服務，並據此決定後續流程。

2. 虛擬角色與語音互動

(1) 利用 Whisper 進行語音辨識，將使用者聲音轉換為文字，使系統能接受自然語音指令並進行語意推理。

(2) 透過 SoVITS 生成自然語音回覆，使系統具備情感式語音輸出能力。

(3) 利用 Unity 3D 建立虛擬 AI 調酒師，能根據指令同步進行動作表現及嘴型、表情呈現，提升系統互動感與沉浸性。



圖 1：Unity 虛擬人物介面

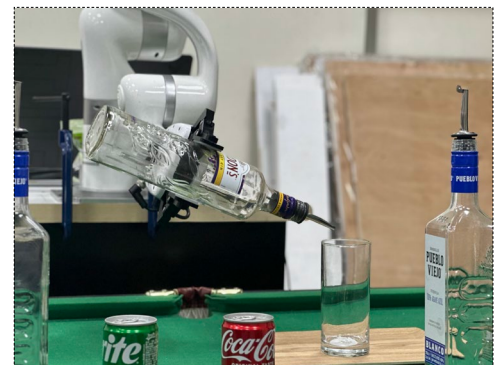


圖 2：xArm 調酒展示

3. 酒瓶辨識與環境感知

採用 YOLO 搭配深度攝影機 (Intel RealSense D435i) 進行酒瓶偵測與定位，並回傳 JSON 格式資料，以輔助調酒流程與狀態確認。

4. 自動調酒行為控制

以 xArm 協作型機械手臂完成調酒步驟，包括抓取酒瓶、倒酒、搖杯與放置酒杯等流程。動作由 LLM 生成結構化指令並傳遞至控制模組執行。

(二)特色

本專題以前饋式大型語言模型為核心，進行語意推論與自主決策，具「自思考」能力。AI 能理解使用者需求、在聊天與任務間自動切換、規劃調酒流程並操控機械手臂執行。結合語音、影像與虛擬角色互動，展示 LLM 在真實場域中串接理解、決策、行動的 Cognitive Robotics 實作。