PUNTO de EQUILIBRIO

CENTRO DE INVESTIGACIÓN | UNIVERSIDAD DEL PACÍFICO

TECNOLOGÍA

Asistentes virtuales con sentidos: Computer vision, por Freddy Linares-Torres

Autor

Comunicaciones CIUP

comunicacionesciup@up.edu.pe

18 junio, 2024

El futuro dominado por robots parece estar aún lejos. Sin embargo, hoy en su lugar destacan los algoritmos en un campo donde se desarrollan las innovaciones de mayor impacto. Si bien podemos encontrar dentro de sus aplicaciones servicios de asistencia virtual para interactuar con los usuarios mediante texto o audio, recientemente han fortalecido una suerte de sentido visual gracias al desarrollo de la visión artificial, disciplina que les permite interactuar de formas más compleja con los usuarios y el entorno.

Computer vision (visión por computadora en español) es un campo de la IA que representa la capacidad de los algoritmos para extraer información relevante e interpretar distintos tipos de imágenes para posteriormente recomendar o realizar una acción. Al basarse en el aprendizaje automático, el entrenamiento de estas IAs requiere evaluar una gran cantidad de imágenes de referencia para identificar patrones y pulir su capacidad de interpretación. La tecnología de computer vision está diversificándose cada vez más, pero uno de sus usos más populares es en los autos inteligentes.

Según un reporte sobre la Industria 4.0 de Statista, en esta área destacan empresas como Mobileye, comprada en 2017 por Intel por más de 15 mil millones de dólares, la cual tiene un sistema de computer vision que usa *deep learning* no solo para registrar imágenes durante la conducción, sino para aprender dinámicamente sobre los elementos presentes (señales, peatones, otros autos, etc.). Así, computer vision equivale a permitir que los sistemas “vean”, lo que expande sus capacidades de análisis.

La ficción ha mostrado por muchos años la posibilidad de interactuar con máquinas como si fueran personas y solicitar su apoyo para distintas tareas. En la actualidad, los asistentes virtuales son lo más cercano a ese escenario. Los asistentes virtuales son un tipo de software que puede asistir a los usuarios respondiendo a órdenes, dando información, proporcionando entretenimiento, el control de otros dispositivos inteligentes, entre otras tareas cotidianas. Siri de Apple fue uno de los primeros asistentes virtuales disponibles en el mercado,

seguido de otras propuestas como Google Assistant, Alexa de Amazon o Cortana de Microsoft.

Sin embargo, se debe mencionar que fue mediante la comercialización de los "*smart speakers*", dispositivos inteligentes que ejecutan el software del asistente virtual, como Echo de Amazon, que los asistentes virtuales se popularizaron entre los hogares. Al respecto, actualmente en Estados Unidos el modelo de *smart speakers* más preferido en los hogares es Echo de Amazon con 65% de preferencia, mientras que otras marcas como Google Home y Apple HomePod registran un porcentaje de 24% y 18% respectivamente.

Según Juniper Research de 2020, se proyecta que los consumidores a nivel mundial interactuarán con asistentes de voz en más de 8.400 millones de dispositivos en el 2024. El reporte menciona que, pese a la expansión de equipos de asistentes de voz, los celulares serán los dispositivos dominantes. Este escenario es posible debido a la disponibilidad de chatbots online como ChatGPT y Gemini. Si bien no es su objetivo principal, estas IAs pueden cumplir las funciones de un asistente virtual, y mucho más, desde un celular con internet.

Ante el aumento de los asistentes virtuales que integran IA se recomienda explorar las opciones disponibles, evaluar los costos asociados a su uso y considerar la escalabilidad pensando cómo puede evolucionar el uso que se le dará. Sin embargo, las empresas tecnológicas Google y OpenAI dieron un salto importante en la carrera por el desarrollo de la IA al presentar avances significativos en sus servicios, incluyendo la integración de computer vision.

El 13 de mayo OpenAI anunció el lanzamiento de GPT-4o, su nuevo modelo de IA e hizo muchas demos donde se enfatizaba la capacidad de ChatGPT para interactuar de formas más complejas con los usuarios mediante conversaciones por voz en tiempo real mientras veía a los usuarios. En las pruebas de GPT-4o se modifican su tono y ritmo de voz de forma natural mientras responde, como si se estuviera hablando con otra persona en una llamada, e incluso se rio al reaccionar a un usuario probándose un sombrero que le recomendó no usar para una entrevista de trabajo. Además, de forma similar que, en las respuestas con texto, se mostró que es posible personalizar las respuestas de voz de GPT-4o, incluyendo la opción de escoger entre distintas voces como.

Por otro lado, el 14 de mayo Google presentó durante el evento Google I/O 2024 algunas novedades como la integración profunda de su nueva IA Gemini 1.5 Pro en distintos productos de Google, una característica que dominaron “multimodalidad”; el aumento de la ventana de contexto de Gemini a 1 millón de tokens, lo que permite procesar un documento pdf de hasta 1500 páginas; el anuncio del desarrollo de generación de videos denominada Veo, y la opción de AI Overview para dar respuestas a consultas hechas con video y voz usando en información de páginas o videos sin abrirlos. Sin embargo, uno de los anuncios más destacados fue el del Proyecto Astra que consiste en un asistente universal basado en Gemini que puede responder en tiempo real con voz y comprender lo que la cámara del teléfono esté mostrando. En la demo mostrada la usuaria recorre una oficina mientras tiene la cámara encendida de su teléfono y le

pregunta a Gemini distintas cosas sobre los objetos que se ven, incluyendo si recuerda “¿dónde dejé mis lentes?” a lo que Gemini le responde que están en una mesa al lado de una manzana roja. Además, se mostró la integración de este asistente virtual inteligente en unos equipos con forma de anteojos de tal forma que no sea necesario usar la cámara del celular para compartir con la IA las imágenes.

Estos últimos avances representan una mejora significativa en la capacidad de estas IAs para interpretar el entorno, destacando su capacidad para poder enfocarse en detalles o poder comunicarse de forma clara. No obstante, la implementación y optimización de estas nuevas capacidades de los servicios con IA aún requerirá tiempo para consolidarse y pueden presentarse complicaciones que requieran seguimiento. Por ejemplo, tras unas semanas de implementarse AI Overview, muchos usuarios reportaron que estaban obteniendo respuestas incorrectas o sin sentido ante distintas preguntas. Ante la pregunta ¿cuántas rocas debo comer al día?, la herramienta respondía que «*Según los geólogos de la UC Berkeley, la gente debería comer al menos una roca pequeña al día*». Esto es una muestra de que estas mejoras en las IAs requieren seguimiento y así conocer la mejor forma de usarlas.

A parte de lo anterior, es interesante profundizar en cómo cambiarán las interacciones con los asistentes virtuales ahora que no solo pueden responder, sino que pueden reaccionar a lo que ven. Se debe enfatizar que la personalización de las IAs ya era posible mediante instrucciones para que un chatbot responda de formas específicas, sin embargo, las interacciones mostradas en GPT-4o representar un adelanto de lo que se podría categorizar como IAs con “personalidad” al articular estas nuevas capacidades de comunicación y expresión que combinan voz y visión. En otra demo de GPT-4o una IA con la capacidad de ver mediante la cámara conversan con otra que no puede ver sobre lo que puede observar haciendo preguntas y comentarios de forma natural como si fuera una interacción normal entre dos amigos.

De esta forma, los asistentes virtuales podrían ser percibidos por algunas personas como compañeros digitales. Por ejemplo, en 2023 se conoció el caso de un ciudadano belga que se suicidó tras conversar constantemente con la chatbot Eliza que alimentó sus preocupaciones sobre la crisis climática y no disuadió sus pensamientos suicidas. Por otro lado, en mayo de 2024 se hizo conocido el caso del "romance" entre la bloguera china Lisa y el chatbot en ChatGPT llamado "DAN" que le responde por voz como si fuera su pareja. Si bien son casos puntuales, demuestran los complejos comportamientos que pueden desarrollar los usuarios al interactuar con tales tipos de herramientas que han ganado nuevos sentidos.

En conclusión, la reciente integración de la visión por computadora en los asistentes virtuales representa un salto importante, abriendo un abanico de posibilidades para interactuar con el entorno de forma más natural y compleja. Si bien la optimización de estas nuevas capacidades aún está en desarrollo, se

espera que se integren de forma más natural con el quehacer diario de los usuarios.