

El módulo de voz de ChatGPT

Autor: Zsolt Lengyel, iTStudy Hungary Training and Research Centre

Revisora: Mária Hartyányi, iTStudy Hungary Training and Research Centre

Objetivo de aprendizaje

Hablar es la forma más natural de comunicarse, y los asistentes de IA ya pueden mantener conversaciones habladas en tiempo real. El objetivo de este material es presentar la interacción con la IA por voz y el papel que puede desempeñar en la enseñanza y el aprendizaje. Al finalizar, los docentes serán capaces de distinguir el dictado de la conversación por voz, describir cómo funciona actualmente ChatGPT Voice, identificar usos pertinentes para la lluvia de ideas, la práctica de idiomas, la accesibilidad, la ayuda profesional rápida y el diálogo reflexivo, y tener en cuenta limitaciones prácticas como los errores de transcripción, el ruido en el aula y la protección de datos al utilizar la voz con el alumnado.

Introducción: dos formas de usar la voz

ChatGPT puede utilizarse por voz de dos formas prácticas. El **dictado** simplemente convierte nuestra voz en texto: hablamos, el sistema transcribe y enviamos el mensaje como si lo hubiéramos escrito. Resulta útil cuando queremos formular rápidamente ideas más largas o cuando escribir es incómodo, por ejemplo, en el móvil.

La **conversación por voz** es un diálogo real: hablamos con la IA y ella responde en voz alta. Esto permite una interacción más rápida y natural, sobre todo para generar ideas, practicar o cuando no queremos estar escribiendo continuamente. La diferencia clave: el dictado convierte la voz en texto, mientras que la conversación por voz es una interfaz conversacional en la que la respuesta de la IA también llega hablada.

Cómo funciona hoy ChatGPT Voice

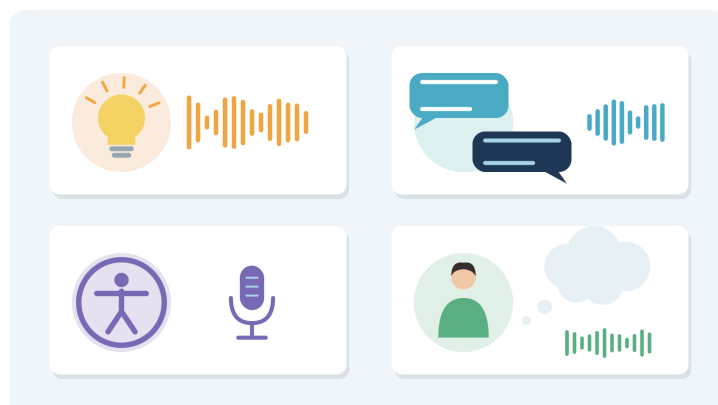
La voz ha evolucionado rápidamente. Desde finales de 2025, las conversaciones por voz tienen lugar directamente dentro del chat, de modo que el intercambio hablado y cualquier contenido visual (por ejemplo, un mapa, una tabla o una imagen) aparecen en la misma conversación. En julio de 2026, OpenAI presentó **GPT-Live**, una nueva generación de modelos de voz capaces de escuchar y hablar al mismo tiempo. Esto hace que los turnos de palabra y las interrupciones resulten naturales: podemos intervenir, el modelo puede indicar que está escuchando y la conversación fluye más como un diálogo entre personas.



Conversación por voz en dúplex completo: el usuario y la IA pueden hablar y escuchar al mismo tiempo

- El **modo Live** (el predeterminado para la mayoría de los usuarios) permite una conversación natural de ida y vuelta y puede utilizar la búsqueda web y la memoria; para razonamientos complejos o tareas de investigación, puede derivar la tarea a un modelo de texto más potente mientras la conversación continúa.
- El **modo Advanced** conserva funciones anteriores en tiempo real, como compartir la cámara o la pantalla en dispositivos móviles, algo útil para mostrar un objeto, un aparato o un problema.
- La voz también puede utilizarse dentro de los **proyectos** (*Projects*), de modo que la conversación hablada puede apoyarse en los archivos y las instrucciones del proyecto.
- Tras la conversación, aparece una **transcripción** en el historial del chat. Puede no coincidir exactamente con lo que se dijo, sobre todo si las voces se solapan, hay ruido de fondo o la conversación es rápida.
- Las sesiones tienen límites de uso que dependen de la suscripción. El idioma del reconocimiento de voz puede configurarse en los ajustes, lo que mejora la precisión.

Posibles funciones en la enseñanza y el aprendizaje



Lluvia de ideas, práctica de idiomas, accesibilidad y diálogo reflexivo: cuatro usos típicos de la interacción por voz

Lluvia de ideas y planificación sobre la marcha

Muchos docentes tienen sus mejores ideas mientras se desplazan o pasean. Una conversación hablada es una forma rápida de pensar en voz alta sobre

una clase: “Mañana enseñó el cambio de neumáticos a aprendices de primer curso. Ayúdame a encontrar un comienzo motivador.” Después, pida un resumen escrito en el chat y edítelo.

Práctica de idiomas

La voz resulta especialmente valiosa para practicar la comunicación profesional en una lengua extranjera: un recepcionista que da la bienvenida a un huésped del hotel, un mecánico que explica una reparación a un cliente, un camarero que toma nota de un pedido. La IA puede hacer de interlocutor, hablar más despacio o más deprisa, repetir y dar retroalimentación sobre el vocabulario. La retroalimentación sobre la pronunciación debe tomarse con cautela; el docente sigue siendo la referencia.

Accesibilidad

La interacción por voz puede reducir las barreras para el alumnado con dificultades de lectura, discapacidad visual o dificultades motoras: puede hacer preguntas y recibir explicaciones sin leer ni escribir. Para este alumnado, la voz puede ser un apoyo real, dentro del marco del centro para las tecnologías de apoyo.

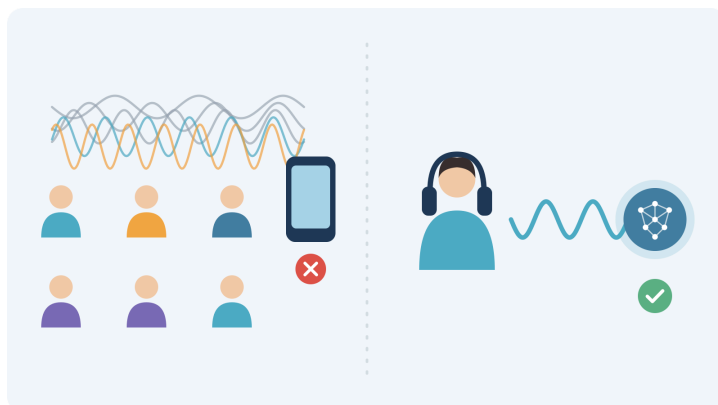
Ayuda profesional rápida

En los talleres y en las prácticas, las manos suelen estar ocupadas. Una pregunta hablada (“¿Qué par de apriete suele indicarse para...?”) puede ser más rápida que escribirla. Sin embargo, la información crítica para la seguridad debe comprobarse siempre en la documentación del fabricante o en la norma correspondiente: una respuesta oral dada con aplomo no es más fiable que una escrita.

Diálogo reflexivo

La conversación por voz se presta bien a la reflexión: repasar una clase que no salió como estaba previsto, prepararse para una conversación difícil con las familias o, en el caso del alumnado, practicar una entrevista de trabajo. Pedir a la IA que asuma el papel de un amigo crítico que hace preguntas, en lugar de uno que simplemente da la razón, hace que la reflexión sea más útil.

Limitaciones prácticas



En un aula ruidosa, el reconocimiento empeora; la voz funciona mejor de forma individual y en un entorno tranquilo

- **Errores de transcripción:** los acentos, las variedades dialectales, la mezcla de idiomas y la terminología especializada pueden reconocerse mal. Revise la transcripción cuando los detalles sean importantes.
- **Ruido:** el ruido de fondo y varias personas hablando a la vez pueden provocar respuestas no deseadas. La voz funciona mejor de forma individual; en clase, utilice auriculares con micrófono o un rincón tranquilo; en el iPhone, el ajuste de aislamiento de voz puede ayudar.
- **Sin fuentes visibles:** una respuesta hablada no muestra sus fuentes con tanta claridad como una escrita. Para los datos, pida las fuentes en el chat y compruébelas.
- **Capacidad de persuasión:** una voz fluida y amable puede hacer que las respuestas erróneas suenen más convincentes. Las alucinaciones no desaparecen en el modo de voz.
- **Límites:** existen límites de duración de las sesiones y de uso, y algunas funciones (p. ej., compartir la pantalla) solo están disponibles en determinados modos.

Protección de datos y uso adecuado con el alumnado



Las grabaciones de voz y las transcripciones son datos personales: hay que saber cómo se almacenan y quién puede acceder a ellas

- En ChatGPT, los fragmentos de audio de las conversaciones por voz se almacenan con la transcripción en el historial del chat y, según el proveedor, se conservan durante 30 días; no se utilizan para el entrenamiento, salvo que el usuario decida compartirlos o active el ajuste correspondiente. Consulte las condiciones vigentes antes de utilizarlo.
- La voz de una persona es un dato personal. No grabe al alumnado ni a sus compañeros sin su consentimiento, y respete el RGPD y las normas de protección de datos del centro.
- Respete los requisitos de edad de los proveedores (en el caso de ChatGPT: al menos 13 años y consentimiento de los padres o tutores para menores de 18) y las normas del centro. Con el alumnado más joven, recurra a demostraciones dirigidas por el docente.
- Acuerde unas normas: nada de información personal o sensible en las conversaciones por voz y una finalidad clara para cada uso.

Ideas clave para docentes

- Distinga el dictado (de voz a texto) de la conversación por voz (diálogo hablado).
- La voz es muy útil para la lluvia de ideas, la práctica de idiomas, la accesibilidad y la reflexión.
- Configure el idioma, busque un entorno tranquilo y revise la transcripción cuando los detalles sean importantes.
- Las respuestas habladas no son más fiables que las escritas: verifique los datos y la información de seguridad.
- Trate la voz como un dato personal: se aplican el consentimiento, los requisitos de edad y las normas del centro.

Referencias

- OpenAI (2026): Introducing GPT-Live. Blog de OpenAI, 8 de julio de 2026.
- OpenAI Help Center (2026): ChatGPT Voice; ChatGPT release notes.
- Lengyel, Zs. (2026): Concrete applications for content generation. Presentación de la formación VITA.

Enlaces

- ChatGPT Voice – preguntas frecuentes:
help.openai.com/en/articles/8400625-voice-mode-faq
- Introducing GPT-Live: openai.com/index/introducing-gpt-live

Derechos de autor © 2026 Consorcio VITA

Esta obra puede compartirse y adaptarse bajo los términos de la Licencia Internacional Creative Commons Atribución/Reconocimiento-CompartirIgual 4.0 (CC BY-SA 4.0).



Financiado por la Unión Europea. Las opiniones expresadas son únicamente las de los autores y no reflejan necesariamente las de la Unión Europea ni las de la Agencia Nacional.