

El flujo completo: de la voz humana a la respuesta hablada

Datos de esta lección

Fase	F1 — Cimientos, seguridad y arquitectura del sistema
Módulo	M1.2 — Modelos mentales: radiografía de un agente telefónico
Puntos de profesionalidad al superar el test	15

Objetivo práctico

Al terminar esta lección podrás describir, paso a paso y sin saltos, el recorrido completo de una llamada desde que una persona empieza a hablar hasta que escucha la respuesta del agente, situando correctamente cada pieza del sistema.

Resultado que obtendrás

Un esquema mental lineal completo: persona habla → STT → LLM → (posible Tool Calling/automatización) → TTS → persona escucha. Este esquema es la herramienta de diagnóstico más importante de todo el curso.

Dónde encaja dentro del sistema final

Esta lección une expresamente todas las piezas anteriores (LLM, STT, TTS) en un único proceso lineal, y añade las dos piezas que faltan para completar el sistema: la automatización (que consulta o actualiza sistemas externos) y el calendario. Es el mapa completo que usarás para diagnosticar durante todo el curso.

Conocimientos previos

- LLM, STT y TTS (las tres lecciones anteriores).

Herramientas necesarias

- Un navegador con NVDA activo, en modo de exploración.

Posible coste

Esta lección es conceptual y no requiere ningún gasto.

Modelo mental

Piensa en una centralita telefónica con un telefonista capaz de escuchar, consultar instrucciones y responder. El proceso completo de atender una llamada, en una oficina tradicional con un telefonista humano, sigue exactamente esta secuencia: alguien llama y habla, el telefonista escucha y entiende lo que le piden, consulta su manual de instrucciones o pregunta a otro departamento si hace falta, decide qué responder, y lo dice en voz alta. El sistema de este curso automatiza cada uno de esos pasos con una pieza tecnológica concreta.

Explicación

El flujo completo de una llamada gestionada por un agente de IA telefónico sigue esta secuencia lineal:

1. **La persona habla.** Su voz llega al sistema a través de la línea telefónica (Twilio, en este curso).
2. **STT (oído):** la voz se convierte en texto.
3. **LLM (cerebro):** el texto se procesa según el Prompt Maestro y el contexto de la conversación, y se decide qué responder.
4. **Tool Calling / automatización (cuando es necesario):** si la conversación requiere consultar o modificar algo externo —por ejemplo, comprobar disponibilidad en un calendario—, el LLM «llama a una herramienta»: se dispara un Webhook hacia n8n o Make, que a su vez habla con Google Calendar o Cal.com, y devuelve una respuesta.
5. **TTS (boca):** la respuesta final en texto se convierte en audio.
6. **La persona escucha la respuesta.**

Este esquema es la base del «Principio de diagnóstico por descarte» que se usa en todo el curso (módulo 4.1 y 4.2): cuando algo falla, la pregunta no es «¿qué ha fallado en general?», sino «¿en qué punto exacto de esta cadena se ha roto la conversación?». Cada fallo tiene una firma distinta según en qué eslabón ocurre.

Vocabulario nuevo

Tool Calling

Piénsalo así: El telefonista levantando el teléfono interno para preguntar a otro departamento antes de responder.

Mecanismo por el que un LLM, en mitad de una conversación, puede solicitar la ejecución de una acción externa (por ejemplo, consultar un calendario) y recibir el resultado antes de continuar respondiendo.

Ejemplo: Cuando alguien pide una cita, el LLM no inventa la disponibilidad: activa una herramienta que consulta el calendario real.

Preparación

No se requiere preparación técnica.

Procedimiento paso a paso

1. **Contexto:** Vas a usar el Buscador del curso, una página real de esta plataforma, para localizar esta lección a partir del término «Tool Calling».

Acción de teclado: Abre el enlace «Buscador» de la navegación principal (o ve a `/buscar.php`). Pulsa `Tab` hasta el campo «Término de búsqueda», escribe `Tool Calling` y pulsa `Enter`.

Respuesta esperada de NVDA: NVDA debería anunciar «Resultados para Tool Calling, encabezado de nivel 2» y, a continuación, el número de resultados encontrados.

Qué significa: El buscador ha encontrado esta lección porque «Tool Calling» aparece en su vocabulario, aunque no esté en el título.

Acción siguiente: Pulsa `Tab` hasta el enlace del resultado y confirma que te lleva de vuelta a esta misma lección.

2. **Contexto:** Comprueba ahora qué ocurre cuando una búsqueda no encuentra ningún resultado, para reconocer ese mensaje si te ocurre más adelante en el curso.

Acción de teclado: Vuelve al campo «Término de búsqueda» y escribe una palabra inventada, por ejemplo `xzqwplft`, y pulsa `Enter`.

Respuesta esperada de NVDA: NVDA debería leer el mensaje «No se han encontrado lecciones que contengan ese término...».

Qué significa: El buscador funciona íntegramente en el servidor (sin JavaScript obligatorio) y siempre da una respuesta textual clara, tanto si hay resultados

como si no.

Acción siguiente: Repite la búsqueda con un término real de este curso, como «Webhook» o «JSON», para seguir explorando.

Errores frecuentes y diagnóstico

Ante cualquier fallo, se revisan todas las piezas del sistema a la vez, sin orden.

Cómo localizarlo: Es ineficiente y genera confusión: la mayoría de fallos tienen una firma que apunta a un único eslabón de la cadena.

Solución: Aplicar el esquema lineal de esta lección: identificar en qué punto exacto se rompe la conversación (¿no entiende lo que se dice? ¿entiende pero responde mal? ¿responde bien pero no ejecuta la acción?) antes de investigar nada.

Comprobación final

Dibuja (en papel, en un documento de texto, o mentalmente en voz alta) el esquema completo: persona habla → STT → LLM → Tool Calling/automatización (si aplica) → TTS → persona escucha. Si puedes recitarlo sin mirar la lección, la comprobación es positiva.

Qué acabas de conseguir

Tienes el mapa completo del sistema que construirás a lo largo del curso, y la herramienta de diagnóstico que usarás constantemente a partir de la Fase 4.

Ejercicio práctico

Para cada uno de estos tres síntomas, indica en qué punto del flujo buscarías primero: (1) el agente no entiende ni una palabra de lo que se dice; (2) el agente entiende perfectamente pero da una respuesta que no tiene sentido; (3) el agente confirma una cita en la llamada pero no aparece en el calendario real.

Resumen de lo imprescindible

- El flujo completo es: persona habla → STT → LLM → (Tool Calling/automatización si aplica) → TTS → persona escucha.
- Cada fallo real tiene una firma que apunta a un eslabón concreto de esta cadena.

- El diagnóstico eficiente empieza por identificar el eslabón, no por revisar todo el sistema a la vez.

No necesitas aprender todavía

Todavía no necesitas saber configurar Webhooks ni automatizaciones: eso se practica en la Fase 3. Aquí basta con entender dónde encajan dentro del flujo.

Estoy preparado para continuar si puedo...

- Recitar de memoria el flujo completo de una llamada.
- Identificar en qué eslabón del flujo buscaría según el síntoma descrito.
- Explicar qué es el Tool Calling y cuándo se activa dentro del flujo.

Test de comprobación

Para registrar el resultado y obtener Puntos de profesionalidad, realiza este test desde la versión web de la lección. Las opciones se muestran aquí sin señalar la respuesta correcta.

1. Una llamada transcurre con normalidad, el agente conversa correctamente y confirma verbalmente una cita, pero esa cita nunca aparece en el calendario real del negocio. Siguiendo el flujo completo de esta lección, ¿qué parte de la cadena tiene sentido investigar primero?
 - a. El STT, porque la transcripción inicial pudo fallar.
 - b. El TTS, porque la voz de confirmación pudo sonar poco clara.
 - c. El Tool Calling y la automatización (Webhook hacia n8n/Make y su conexión con el calendario), porque la conversación en sí ya ha funcionado correctamente.
2. Una persona llama y, desde el primer segundo, el agente responde cosas que no tienen relación con lo que se ha dicho. Aplicando el esquema lineal de esta lección, ¿en qué tramo del flujo tiene más sentido buscar primero?
 - a. Al principio del flujo (STT), porque el problema aparece desde la primera respuesta, antes de que intervenga ninguna automatización.
 - b. En el calendario, porque cualquier fallo de conversación se origina siempre ahí.
 - c. En el Tool Calling, porque es la pieza más compleja del sistema.
3. ¿Por qué el esquema TELÉFONO → AGENTE → CEREBRO → VOZ → AUTOMATIZACIÓN → CALENDARIO se describe en esta lección como «la

herramienta de diagnóstico más importante del curso»?

- a. Porque permite localizar en qué eslabón concreto se rompe una conversación, en lugar de revisar todo el sistema a la vez sin ningún orden.
- b. Porque es el único esquema que hay que memorizar para aprobar el curso.
- c. Porque sustituye a la necesidad de revisar los registros (logs) de cada plataforma.

Fuentes

- [Voice agents — OpenAI \(consultado el 2026-08-30\)](#)

Última actualización de este contenido: **2026-02-10**.